Assessing narrative comprehension in young children

ALISON H. PARIS
SCOTT G. PARIS
University of Michigan, Ann Arbor, Michigan, USA

The national commitment to improve reading achievement for all children in the United States has led to many educational reforms in literacy education. Higher standards, early assessment, direct instruction of basic skills, and early intervention to improve reading are all typical in elementary schools today. There is no quick fix for the myriad of difficulties that children encounter as they learn to read, but reading educators have made considerable progress in identifying and correcting children's early difficulties (Allington & Walmsley, 1995; Snow, Burns, & Griffin, 1998). Many of the recent interventions and instructional programs for reading are based on direct instruction of enabling skills such as phonological awareness and decoding so that beginning readers can identify, sound out, or figure out words in text.

During the past 10 years of renewed emphases on beginning reading, less attention has been given to children's comprehension skills than to decoding skills (RAND Reading Study Group, 2001). More research on young children's comprehension skills and strategies is needed in order to diagnose and address children's early reading difficulties that extend beyond decoding. The focus of this article is on children's comprehension of narrative stories and, more specifically, narratives that are illustrated in wordless picture books. We created and tested assessment materials and procedures that can be used with young children, whether or not they can decode print. Such early assessments of comprehension skills can complement existing assessments of enabling skills, provide diagnostic assessments of comprehension problems, and link comprehension assessment with classroom instruction. The rationales for the approach, methods, materials, and age range of children tested are provided in the following sections.

ABSTRACTS

Assessing narrative comprehension in young children

THIS ARTICLE explains the creation and validation of the Narrative Comprehension of Picture Books task (NC task), an assessment of young children's comprehension of wordless picture books. Study 1 explored developmental changes among 158 K–2 children in narrative comprehension and the correlations among children's performance on the NC task and other measures of early reading. There was significant improvement with increasing age on NC task measures. Significant concurrent validity was found between the NC task and oral reading comprehension for readers and between the NC task and several prereading skills for prereaders. Study 2 tested the generalizability of the NC task by giving a sub-sample of students ($n = 91$) two additional picture books using the NC task procedures. Intertask correlations showed that children were consistent on each of the NC task dependent variables across the three books. The same developmental trends by grade and reading ability were evident on all three versions of the task. Study 3 confirmed the generalizability of the NC task across children, books, and testers, and it revealed sensitivity to longitudinal growth in children's comprehension skills. There was little evidence of practice effects influencing NC scores. Study 3 also revealed significant concurrent and predictive validity between NC measures and other assessments of early reading skills such as the Gates McGinitie Reading Test and the Iowa Test of Basic Skills. The NC task is a valid quantitative measure of young children's comprehension that is sensitive to developmental changes and adaptable to various books. We discuss how narrative comprehension is fundamental to beginning reading and how the NC task may be used for classroom instruction and assessment.

Evaluación de la comprensión de narrativas en niños pequeños

ESTE TRABAJO explica la creación y validación de la tarea Comprensión Narrativa de Libros de Imágenes (Narrative Comprehension of Picture Books, tarea NC), un instrumento de evaluación de la comprensión de libros de imágenes, sin palabras, en niños pequeños. El Estudio 1 exploró los cambios evolutivos en la comprensión narrativa en el desempeño de 158 niños de K–2, así como las correlaciones entre el desempeño de los niños en la tarea NC y otras medidas de lectura temprana. Se registró un avance significativo en el desempeño en la tarea NC con el aumento de la edad. Se halló validez concurrente significativa entre la tarea NC y la comprensión en la lectura oral, para los lectores, y entre la tarea NC y varias habilidades previas a la lectura entre los prelectores. El Estudio 2 investgó la generalizabilidad de la tarea NC mediante la evaluación de un subgrupo de estudiantes ($n = 91$) usando los procedimientos de la tarea NC con dos libros de imágenes adicionales. Las correlaciones intratarea mostraron que los niños fueron consistentes en cada una de las variables dependientes de la tarea NC en los tres libros. Se evidenciaron los mismos efectos evolutivos por grado y habilidad de lectura en las tres versiones de la tarea. El Estudio 3 confirmó la generalizabilidad de la tarea NC a través de los niños, los libros y los evaluadores y reveló sensibilidad al crecimiento evolutivo en las habilidades de comprensión de los niños. Se registró poca evidencia de efectos de la práctica que afectasen el desempeño en NC. El Estudio 3 también reveló validez concurrente significativa y validez predictiva entre las medidas NC y otros instrumentos de evaluación de habilidades de lectura temprana, tales como el Test de Lectura Gates McGinitie (Gates McGinitie Reading Test) y el Test Iowa de Habilidades Básicas (Iowa Test of Basic Skills). La tarea NC es una medida cuantitativa válida de la comprensión en niños pequeños, sensible a los cambios evolutivos y adaptable a libros diversos. Discutimos de qué modo la comprensión narrativa es fundamental para los lectores iniciales y cómo puede usarse la tarea NC para la enseñanza y evaluación en el aula.

Bewerten des Erzählungsverständnisses in jungen Kindern

DIES ARTIKEL erklärt das Hervorbringen und Bewerten des Erzählungsverständnisses von Bilderbuchschularbeit (Narrative Comprehension Task bzw. NC-Task), als eine Verständnisbewertung junger Kinder von Bilderbüchern ohne Worte. Die Studie 1 untersuchte Entwicklungsveränderungen unter 158 K–2 Kindern im erzählerischen Verständnis und in den Wechselbeziehungen zwischen der vorstellerischen Leistung der Kinder bei den NC-Schularbeiten und bei sonstigen Bewertungen des frühen Lesens. Es gab eine erhebliche Verbesserung mit steigendem Alter bei den NC-Aufgabenbewertungen. Bezeichnend-übereinstimmende Werte fanden sich zwischen NC-Aufgaben und mündlichem Leseverständnis bei Lesern und zwischen NC-Aufgaben und verschiedenen Anfangsübungen bei Leseanfängern. Studie 2 prüfte die Allgemeingültigkeit der NC-Aufgaben durch Ausgabe eines Musterbeispiels an Schülern ($n = 91$) von zwei zusätzlichen Bilderbüchern unter Anwendung der NC-Aufgabenbewertungen. Zwischenwertkorrelationen zeigten, dass die Kinder sich bei jeder der von den NC-Aufgaben abgeleiteten Variablen gegenüber allen drei Büchern im Einklang befanden. Die gleichen Entwicklungstrends, sowohl bei Zensuren und Lesebefähigung, waren bei allen drei Versionen der Aufgaben offensichtlich. Studie 3 bestätigte die Allgemeingültigkeit der NC-Bewertung gegenüber den Kindern, den Büchern und den Prüfern, und offenbarte eine Sensibilität hin zu fortgesetztem Wachstum an Kenntnissen zum besseren Verständnis bei den Kindern. Übungseffekte waren von geringer Bedeutung in der Beeinflussung der NC-Bewertungen. Studie 3 brachte auch eine deutlich-übereinstimmende, vorhersagbare Validität zwischen NC-Werten und anderen Bewertungen früher Lesekenntnisse, wie des McGinitie Lesetests und des Iowa-Tests der Grundkenntnisse. Die NC-Prüfaufgabe ist ein wohlbegründetes, quantitatives Maß zur Einsichtnahme in das Verständnis junger Kinder, welches sensitiv zu entwicklungsbedingten Veränderungen ist und auf unterschiedliche Bücher angewandt werden kann. Wir diskutieren wie fundamental wichtig das Darstellungsverständnis zum Beginn des Lesens ist, und wie die NC-Aufgaben für Anweisung und Bewertung im Klassenzimmer angewandt werden.

幼い子供達の物語理解の評価

本稿では、絵本タスクを用いた物語理解(NCタスク)の創作とその正当性の立証、幼い子供達による文字のない絵本の内容理解の評価を説明する。研究１は、幼稚園児から小学２年生１５８人の子供達の物語理解における発達的変化、また子供達のNCタスクにおけるパフォーマンスと早期リーディング力の他の測定結果との相関を調べた。年齢が大きくなるにつれ、NCタスク測定結果において有意な進歩が見られた。NCタスクと文字を読める子供達にとっての音読理解との間、また、NCタスクと読めるようになる前の子供達にとってのいくつかのプレリーディングスキルとの間に、有意な併存的妥当性が見られた。研究２では、 サブサンプルの生徒(n=91)にNCタスクの方法を用いた絵本をもう２冊与えることで、NCタスクの一般化可能性を調べた。タスク間の相関によって、子供達が本３冊に渡るNCタスク従属変数のそれぞれにおいて一貫していることが示された。学年及びリーディング力ごとの同じ発達上の傾向が、タスクの３つの型すべてにおいて明らかであった。研究３は、子供、本、試験官におけるNCタスクの一般化可能性を確認し、子供の理解技能の長期に渡る成長に感度を示した。練習効果がNCスコアに影響を及ぼした形跡は、ほとんど見られなかった。また、研究３では、NC測定結果と Gates McGinitie リーディングテストやアイオワベーシック技能テストといった早期リーディング技能の評価との間に有意な併存的妥当性及び予測的妥当性が示された。NCタスクは、発達変化に対する感度がよく、様々な本に適応可能で、幼い子供達の理解を計るために妥当な量的測定法である。ここでは、物語理解が初期のリーディングにとっていかに重要であり、 NCタスクが教室における指導や評価でいかに使えるのかを検討する。

Évaluation de la compréhension du récit chez de jeunes enfants

CE TEXTE explique la construction et la validation d'une épreuve de Compréhension du Récit de livres d'images (tâche CR), une évaluation de la compréhension par de jeunes enfants de livres illustrés mais sans texte. La première étude a exploré les changements dans le temps de 158 enfants de 2ᵉ année de jardin d'enfants en compréhension de récit et les corrélations entre les performances des enfants dans la tâche CR et d'autres mesures de début de lecture. On a observé une progression significative avec l'âge dans les mesures de la tâche CR. On a trouvé une validité concurrente significative entre la tâche de CR et la compréhension orale du récit chez les lecteurs, et entre la tâche de CR et plusieurs compétences de prélecture chez les prélecteurs. La deuxième étude a testé la généralisabilité de la tâche de CR en donnant deux livres sans texte supplémentaires avec tâche de CR à un sousé chantillon d'élèves (*n*= 91). Les corrélations intertâches ont montré que les enfants sont fidèles pour les trois livres et pour chacune des variables dépendantes de la tâche de CR. Les mêmes tendances dans le développement sont apparues concernant le niveau scolaire et la capacité de lecture pour les trois versions de la tâche. La troisième étude a confirmé la généralisabilité de la tâche de CR pour les enfants, les livres, et les testeurs, et s'est révélée sensible au développement dans le temps du niveau de compréhension de lecture des enfants. On n'a guère trouvé d'effets de la pratique sur les résultats de CR. La troisième étude a aussi mis en évidence une validité prédictive concurrente significative entre les mesures de CR et d'autres évaluations du niveau de début de lecture tels que le test de lecture de McGinitie et le test des compétences de base de l'Iowa. La tâche de CR est donc une mesure quantitative valide de la compréhension des jeunes enfants, sensible aux changements dans le temps et adaptable à différents livres. La discussion porte sur le caractère fondamental de la compréhension de lecture et sur la façon dont il est possible d'utiliser la tâche de CR pour l'enseignement et l'évaluation de la lecture.

В статье описаны создание и апробация заданий на понимание сюжета иллюстрированных книг (задания NC), которые позволяют оценить, как маленькие дети воспринимают сюжет в бестекстовых книгах. Предметом исследования 1 являлась динамика развития 158 детей 5-8 лет, связанная с пониманием сюжета, и соотношение этих результатов с другими показателями, по которым оценивается чтение детей раннего возраста. Отмечено существенное улучшение результатов у более старших детей. Выявлена прямая связь между пониманием сюжета бестекстовых книг и пониманием текстовых книг, прочитанных вслух, а также между выполнением заданий NC и другими навыками, необходимыми будущим читателям. В исследовании 2 проверялась генерализованость задания NC. Подвыборке учащихся (n = 91) было предложено выполнить NC задания к двум дополнительным иллюстрированным книгам. Выявлено, что при работе с каждой из трех книг дети демонстрировали устойчивые показатели по каждой из NC переменных. Одинаковые тенденции развития, связанные с возрастом и способностью к чтению, были очевидны при выполнении всех трех версий задания. Исследование 3 подтвердило генерализованость задания NC для детей, книг и тестирующих, а также продемонстрировало развитие навыков понимания с возрастом. Выявлено мало подтверждений связи результатов NC с натренированностью детей. Исследование 3 также показало значимую существующую и прогнозируемую взаимосвязь между заданиями NC и другими способами оценки ранних навыков чтения, такими как читательский тест Гейтса МакДжинити и Тест основных навыков, принятый в штате Айова. Задание NC – надежный инструмент для количественного измерения развитости навыков понимания, чувствительный к динамике детского взросления и к различным книгам. Мы показываем, насколько фундаментально понимание сюжета для того, чтобы дети начали читать, и как задание NC может использоваться для обучения и оценки.

Why focus on comprehension?

Recent and renewed emphases on reading comprehension reveal the paucity of theoretical approaches during the past 10 years, perhaps because of the dominant focus on oral language and basic skills. Although it is not our intention to propose a theory of comprehension in this article, it is important to provide a conceptual basis for our research. Our views of meaning making, whether with text, picture sequences, or objects, are based on a cognitive schematic approach that has roots in Piagetian theory as well as schema-driven reading theories (e.g., Anderson & Pearson, 1984) because we emphasize how new information is assimilated into previous understanding. The implications for this line of research are that children view pictures in books and construct meaning according to (a) their ongoing construction of a sensible story schema from the pictures and (b) their previous experiences with similar pictures, schemata, and stories. Additionally, comprehension requires the application of a variety of strategic processes (Paris, Wasik, & Turner, 1991; Pressley et al., 1994) that include making inferences, identifying main ideas, summarizing, predicting, monitoring, and backtracking.

In the present research, children must "read" and elaborate the meanings identified in individual pictures. They must integrate meanings across pictures, delete extraneous pictorial information, retain main ideas, form expectations regarding possible future pictures, monitor ongoing understanding of pictorial information, and backtrack to previous pictures when comprehension fails. We think that comprehension depends on "effort after meaning" and "clicks of comprehension" (Bransford & McCarrell, 1974) that denote how people construct relational meanings of events. When children look at pictures in books, the process of meaning making is similar to the cognitive efforts to construct meaning from printed words. For example, children construct relations among characters, actions, and events in pictures or text based on the contextual clues and their prior knowledge. The meanings are the products of efforts to integrate and reconcile the individual pictures, perhaps syncretically, thematically, or functionally, but all are efforts to construct relations among events in a coherent story. These processes of comprehension may be influenced by vocabulary and language development to some extent, but according to Bransford and McCarrell (1974), "Comprehension results only when the comprehender has sufficient alinguistic information to use the cues specified in linguistic input to create some semantic content that allows him to understand" (p.

204). In a parallel fashion, the narrative comprehension (NC) task assesses children's reasoning about narrative elements and episodes shown in pictures when linguistic cues to meaning are minimized.

Why focus on comprehension of narrative stories?

Narrative stories surround children from their earliest language experiences (Dickenson & Snow, 1987; Heath, 1986; Nelson, 1973; Stein & Albro, 1996). Children as young as 2 to 3 years old develop a rich repertoire of knowledge about narrative as a result of an innate desire to promote their own well-being (Stein & Albro). Children use this knowledge, which is sometimes called "goal-structured knowledge," to narrate their needs, desires, plans, and frustrations; they also use this knowledge to understand and respond to others' demands, requests, needs, and emotional reactions. Manifestations of goal-seeking in language occur even prior to 2 to 3 years of age, for children's earliest words often serve the function of referring to things children want or activities that interest them.

Beyond these rudimentary experiences with narrative stories, young children become increasingly skilled at understanding and producing complex narrative stories (Dickenson & Snow, 1987; Dickenson & Tabors, 1991; Snow, 1983). Some of these stories are first-hand narratives about daily events; some stories are retellings of previous experience; some stories are retold favorite tales; and some are oral fantasy narratives with characters, conflicts, and coherent actions sequences (Botvin & Sutton-Smith, 1977; Heath, 1986; Miller, Wiley, Fung, & Liang, 1997). Narratives become important for communication between adults and children because they are interwoven through so many facets of daily experience. Parents model the construction of narrative for their children by telling personal and family stories in their presence (Fiese et al., 1999; McCabe & Peterson, 1991; Miller, 1994). Parents also conarrate events with their children by providing scaffolding with questions and assisting them in "learning how to narrativize" (Pressley, 1996; Wiley, Rose, Burger, & Miller, 1998). Additionally, stories surround children in their thematic and symbolic play (Pelligrini, 1985). Some stories are based on joint book-reading experiences where children learn to understand and produce narrative as they talk with their parents about characters, actions, intentions, and endings (Dickenson & Smith, 1994; Flood, 1977; Morrow, 1985; Roser & Martinez, 1985; Teale, 1986). The basic narrative structure of charac-

ters involved in goal-directed actions within settings is evident in all these kinds of stories.

Narrative thinking is pervasive in children's early cognitive development. Preschoolers initially construct primitive accounts of story schemas called *narrative scripts*. Narrative scripts are knowledge structures that describe events and sequences of routine activities, such as attending birthday parties and visiting the doctor (Fivush & Hammond, 1990; Nelson & Hudson, 1988). Children then develop narrative schemas, which include knowledge about main events of stories (e.g., characters, settings, problems) as well as concepts about the temporal and causal sequencing of events in stories (Mandler & Johnson, 1977; Stein & Glenn, 1979; Stein & Trabasso, 1982; Yussen & Ozcan, 1996). Preschoolers also acquire understandings of characters' internal responses, including their mental processes, states, and experiences (Astington, 1993; Frye & Moore, 1991; Wellman, 1990). This knowledge is critical to understanding narrative stories. Young children acquire narrative thinking skills, such as representational abilities and "perspective taking" as well as conceptual and working memory capacities, to understand external and internal features of narrative stories (Case, 1992; Case & Okamoto, 1996). In a more general sense, the "narrative mode" is a cognitive foundation by which young children "learn how to mean" in order to make sense of their world (Bruner, 1986).

Narrative comprehension is a complex meaning-making process that depends on the simultaneous development of many skills including, for example, story grammar, theory of mind, and perspective-taking skills. We chose to build an assessment task on narrative comprehension for several reasons.

1. Narrative competence is among the fundamental cognitive skills that influence early reading development. Whitehurst and Lonigan (1998) referred to these skills as "outside-in" skills because children use the semantic, conceptual, and narrative relations that they know to comprehend the text. They may be conceptualized as "top-down" skills by others and generally refer to conceptual knowledge that facilitates comprehension of words and text. In this view, narrative competence is a fundamental aspect of children's comprehension of experiences before they begin to read, and it helps children map their understanding onto texts. The importance and early development of narrative thinking may be one reason that elementary classrooms are dominated by texts in narrative genre (Duke, 2000).

2. Because of the extensive research on narrative comprehension, there is ample documentation of the importance of narrative comprehension among older children and adults, as well as extensive research on developmental aspects of narrative comprehension (e.g., Berman & Slobin, 1994).

3. The clear structure of narrative stories with specific elements and relations provides a basis for assessment of understanding.

4. Narrative is closely connected to many concurrent developmental accomplishments of young children in areas such as language, play, storytelling, television viewing, memory, listening, and reading. It is an authentic experience in young children's lives, and it reveals important cognitive accomplishments.

Why is it important to assess young children's comprehension of pictorial narratives?

We think it is crucial to assess young children's comprehension through listening and viewing in order to minimize the constraints of decoding that may hinder typical assessments of reading comprehension in beginning readers. Wordless picture books with coherent stories permit assessments of comprehension independent of decoding skills. Narrative stories can be communicated verbally, in print, or in pictures, and comprehension can be assessed in the corresponding modalities through listening comprehension, reading comprehension, or picture comprehension. We chose to use pictorial stories because they are familiar to children and fun to look at, are commonly presented in books, are developmental bridges to text with illustrations, and are authentic activities in the lives of children at school and home. Pictorial narratives also require narrative thinking skills similar to the cognitive demands of text-based stories, such as integration of information, inferential skills, knowledge about main story elements, and understanding of temporal and causal sequences (Bornens, 1990; Graham, 1990; Snow & Ninio, 1986). Previous research has shown that from 4 to 8 years of age, the ability to integrate meaning among pictures increases (Bornens, 1990; Paris & Mahoney, 1974), the ability to reason flexibly about picture sequences improves (Schmidt & Paris, 1978), and the ability to make inferences about pictures increases (Schmidt, Paris, & Stober, 1979). Thus, children from preschool through elementary school years are learning to identify implicit and explicit relations among pictures and to integrate the meaning of the pictures. Given the extensive research on narrative competence, it is surprising that there are no cognitive assessments of children's narrative comprehension

that are relevant for reading development. This research was designed to fill that void.

Why is an assessment of narrative comprehension important for young children?

Most parents and reading educators agree that the primary goal of reading is comprehension, and most agree that comprehension is difficult or impossible if the words are not decoded or understood. Because decoding enables comprehension, it has often been taught first to young children and has been regarded as a precursor to reading achievement. The implicit sequence of decoding first and comprehending second is evident in several conceptual approaches to reading. For example, Chall's (1996) stages of reading development emphasize that children first learn to read and then read to learn. The subtle message is that assessment and instruction of comprehension can wait until children demonstrate fluent oral reading, which, according to Chall's model, generally begins at grade 4 (stage 3 of the model). Information processing models of reading comprehension (e.g., LaBerge & Samuels, 1974) assert that comprehension requires mental effort and capacity that are possible only after decoding skills have become automated and, thus, free mental resources for understanding. We agree that automated decoding skills enable better comprehension with increasing age and skill, but we think that the dependency does not imply that comprehension is only derived from decoding nor that comprehension is unimportant for young children.

The prevailing view that comprehension follows decoding is sometimes referred to as a "bottom-up" processing approach. It overshadows comprehension by ignoring how cognitive processes such as schema activation, context, strategy use, and inference are involved in early reading. Indeed, the "reading wars" might be paraphrased as a fight over whether reading words accurately promotes understanding or understanding promotes reading words correctly. The various compromises to the battle emphasize balanced, comprehensive, or interactive approaches that justify the reciprocal influences, rather than strictly sequential influences, between decoding and comprehension. However, balanced approaches require assessments of comprehension that minimize the influence of decoding, just as there are assessments of decoding and word identification that minimize comprehension. That is exactly why assessments of pictorial narratives are important. They allow teachers to assess how children understand narrative stories and their elements and

relations as well as how children analyze, infer, and summarize event sequences—all independent of the ability to decode printed words. We believe that assessments of pictorial narratives minimize the confounding of decoding skills and provide assessments of cognitive processes and verbal abilities that are important for children's early reading. Our research provides evidence of the concurrent and predictive validity of narrative comprehension that supports this claim.

Assessments of children's narrative comprehension can provide useful evidence to teachers. First, the assessments allow teachers to gain insights about children's familiarity with picture books, their book-handling skills, and their responses to books, both affective and cognitive. Second, assessments of narrative stories can be closely connected with classroom instruction in primary grades such as storyboards, puppet play, storytelling, dictated stories, and joint book reading. Similar activities can be reinforced at home between parents and children. Third, assessments that have a uniform set of materials and procedures can yield quantitative information that can be used to measure and report individual growth and progress. With increasing pressures to document young children's literacy progress in primary grades, such instruments are needed to complement traditional informal assessments based on observations of oral reading accuracy. Fourth, narrative assessments allow multiple measures of children's reading skills to be collected in the same testing episode. Current research provides several discrete measures of children's Picture Walk behaviors (see Appendix A for a description of Picture Walk activities), retellings of stories, and comprehension of implicit and explicit information in response to questions. The measures are diverse and more appropriate for young children than filling in answer bubbles or worksheets. The multiple measures and authenticity of the assessment procedures make them child centered and teacher friendly.

Preview of the three studies

The purpose of this research was to provide a uniform task and procedure for assessing children's narrative comprehension independent of decoding abilities and to evaluate the reliability and validity of the assessment. We believe that children's narrative thinking is a fundamental contributor to early reading comprehension and that assessments of narrative thinking with pictures can identify children's comprehension strengths and difficulties. Such a comprehension focus in early reading assessment and

instruction complements the traditional focus on ba-sic decoding skills. The first study in this article de-scribes the narrative comprehension (NC) assessment task and provides evidence about devel-opmental patterns of performance on the NC task and its relations to other measures of early reading. Study 2 demonstrates the generalizabilty of the NC task across three different picture books and shows that children score similarly on the Picture Walks, retellings, and comprehension questions for all three books. Study 3 shows compelling evidence of the re-liability of the task and individual growth after one year as well as evidence of concurrent and predictive validity.

Study 1

Many wordless picture books for young chil-dren do not provide stories in a narrative genre. Instead, they show pictures related to themes (e.g., animals, transportation); they illustrate sequences (e.g., numbers, the alphabet); or they provide predictable schemata (e.g., *Good Night Moon* by Margaret Wise Brown) as ways to connect young children's knowledge to book illustrations. Text often accompanies narratives for children in familiar folk tales or novel stories, and it is customary for adults to read the text while children look at the pictures. These typical books do not allow assessment of nar-rative comprehension because they either (a) fail to provide coherent narratives in pictures or (b) con-found picture comprehension with text. Likewise, there are limited instruments available for assessing narrative comprehension of picture books.

Morrow (1990) assessed children's "construc-tion of narrative" by observing child-adult discus-sions while reading, by eliciting a retelling, and by categorizing each child's behaviors according to where the focus was placed (e.g., on print, story structure, illustrations, questions, comments). Although Morrow identified this process as a mea-sure of narrative meaning making, it is really more about the social process of meaning making and the interactions that occur among adult, child, and book. This measure does not assess the quality of children's comprehension, and it is not specific to narrative thinking that requires integration of genre-specific information. Similarly, Sulzby (1985) de-scribed how children's language and text reading display a developmental progression—from viewing individual storybook pages as if they were discrete units to treating the pictures as parts of an integrated whole. It is an assessment of the degree that chil-dren's language and behavior while reading texts ap-proximate conventional reading. Sulzby differentiat-ed between page- and story-level descriptions in order to assess children's abilities to create a coherent story across pictures. Sulzby's assessment has provid-ed the field with valuable measures of story language and storytelling skills, but her assessment does not assess children's comprehension of specific story in-formation.

Van Kraayenoord and Paris (1996) created an assessment called "Story Construction from a Picture Book" that measured Australian children's abilities to construct meaning from pictures. Six aspects of meaning making were measured, including initial examination of the book, remarks about the pictures, elaboration, metalinguistics, revision strategies, and identification of themes or morals. The authors found that children's abilities to construct stories from picture books at 5 to 6 years of age were corre-lated significantly with standardized reading test scores two years later.

In a procedure similar to that used by van Kraayenoord and Paris (1996), we modified a trade book with a clear narrative story line—a strategy that can be used easily for both assessment and instruc-tional purposes. We located commercially published wordless picture books, adapted them by deleting some irrelevant pages to shorten the task, and assem-bled the pages of photocopied black-and-white pic-tures into spiral-bound little books. Selection of the picture books depended on several criteria that were necessary to make them "good" wordless picture books for use with the NC task. Most important, the pictures of the book had to tell the story themselves without depending on text. It was also important that the story line revealed by the pictures was clear with an obvious sequence of events and that the pic-tures contained the main elements of stories (i.e., settings, characters, problems, resolutions). In our search for appropriate picture books, we discovered that most picture books do not use vibrant colorful pictures that tell a clear story and that picture books with simple black line drawings were much more likely to meet our criteria for a good book. Therefore, all four of the picture books that we used in the three studies of our research contained simple black-and-white illustrations. Furthermore, while materials of any length would have sufficed, we chose to delete a few pages in order to help simplify the story line and ensure that the sequence of main events was very clear. Despite these superficial changes, the NC task involved authentic-looking books.

The first study reports the NC task procedures for observing how children interacted with wordless picture books under three conditions: spontaneous examination during a Picture Walk, elicited retelling, and prompted comprehension during questioning. The Picture Walk procedure allows children to become familiar with the story before being questioned—a practice recommended for young children (Fountas & Pinnell, 1996). The retelling phase provides a measure of the child's free recall of the story. This retelling differs from those typically elicited from prereaders (e.g., Morrow, 1990) because children retell a story after "reading" pictures rather than after listening to a story read orally by an adult. Similar to the cognitive demands of retelling a story after reading text, the present retelling method requires children to read the pictures strategically and to construct meaning actively in order to retell a complete and integrated story. The prompted comprehension phase provides a uniform and quantified procedure for eliciting and scoring children's understanding of narrative elements and relations. We administered the NC task to children in grades K–2 in order to assess the developmental appropriateness of the testing procedures and the measurement sensitivity of the task.

Additional data about the children's reading performances were collected in order to determine how performance on the NC task is correlated with other emerging reading skills. Leslie and Caldwell's (1995) Qualitative Reading Inventory II (QRI-II) was administered to children individually and provided diagnostic information about children's word identification, oral reading accuracy, and comprehension and memory of text. The QRI-II includes narrative and expository passages arranged by grade level and difficulty. In addition, we collected data from several tasks included in the Michigan Literacy Progress Profile (MLPP), which included phonemic awareness (PA), hearing and recording sounds (HRS), and concepts about print (CAP). The multiple tasks allowed us to calculate the correlations between the NC task and QRI-II measures for children who could read and NC performance and three MLPP enabling skills for prereaders. These correlations permit tests of the concurrent validity, as well as degree of independence, of the various skills.

Method

Subjects

One hundred fifty-eight K–2 students from one elementary school in a Michigan city participat-

TABLE 1

NUMBERS OF PARTICIPANTS BY STUDY, GRADE, AND READER STATUS

	Study 1 (n = 158)	Study 2 (n = 91)	Study 3 (n = 141)
Kindergarten			
Prereader	33	31	45
Reader	1	0	7
Total	34	31	52
First grade			
Prereader	36	35	9
Reader	25	11	39
Total	61	46	48
Second grade			
Prereader	4	4	0
Reader	59	10	41
Total	63	14	41

ed in the study during the fall of 1998. The school serves families with diverse socioeconomic backgrounds. Students were randomly selected from among those who returned permission letters in 14 classrooms. Their ages ranged from 61 to 99 months ($M = 81$, $SD = 10$). There were approximately equal numbers of females ($n = 81$) and males ($n = 77$), and the sample was ethnically diverse, with 49% Caucasian, 22% African American, 12% Asian American, and 14% other or multiracial. As shown in Table 1, almost half of the students were prereaders at the start of the study; most kindergartners were prereaders; and most second-grade students were readers.

Task

The first book used in the NC assessment was *Robot-Bot-Bot* by Fernando Krahn (1979, Dutton), which tells the story of a family whose new robot "housecleaner" goes wild after the child plays with its wires. The book is illustrated in black line drawings with no accompanying text and has a clear story line with an obvious sequence of events and main elements of stories. The adapted version of the robot book omitted a few pictures. The remaining 18 pages were photocopied and assembled into a book format with a spiral binding and cover. The title and author's name printed on the front cover were the only words in the book. The task has three parts (Picture Walk, retelling, prompted comprehension) that yield five different scores: (a) spontaneous reactions to the story, (b) retelling of the story, (c) comprehension of explicit story information, (d)

comprehension of implicit story information, and (e) total storybook comprehension.

Part 1: *Storybook Picture Walk*. The Picture Walk allows observations of children's spontaneous and independent interactions while reading the picture book. Children were first given a closed book and asked to look through it. They were then encouraged to say out loud whatever they were thinking about as they looked at the pictures. While children read the story, observations were made about the following five types of Picture Walk elements: book-handling skills, engagement behaviors, picture comments, storytelling comments, and comprehension strategies. (See Appendix A for a complete description.) These five types of behaviors were derived from a pilot study, wherein the behaviors that children displayed while independently viewing a picture book were documented and categories created to represent clusters of behaviors.

Book-handling skills were defined to assess how children oriented the book, whether they had a sense of appropriate speed and order for turning pages, and whether pages were skipped or skimmed. *Engagement* was defined as behavioral and emotional involvement and judged by attention, interest, affect, and effort. *Picture comments* described whether the child made discrete comments about a picture, including descriptions of objects, characters, emotions, actions, or opinions, as well as character vocalizations. *Storytelling comments* were defined as integrative comments among pictures, which demonstrated an understanding that pictures are related in a larger story. This might include narration, dialogue, and the use of storytelling voice or language. *Comprehension strategies* evaluated whether the child displayed vocalizations or behaviors that showed attempts at comprehension such as self-correction of story elements or narrative, looking back or ahead in the book in order to aid in creating the narrative, asking questions for understanding, and making predictions about the story.

In a righthand column adjacent to each of these elements, a 0–2-point scoring rubric described the behaviors appropriate for a 0-, 1-, or 2-point score, and they represented the depth of spontaneous reactions. For example, children received 0 points for storytelling comments if they gave no verbalizations about the pictures. They received 1 point for inconsistent and discrete provision of storytelling elements and 2 points for storytelling comments that connected the events of the story through dialogue or narration. The experimenter made the 0–2-point judgments on each element during the Picture Walk, and children received a total Picture Walk score that could range from 0 to 10, with higher scores indicating higher levels of spontaneous, appropriate interactions with the book.

Part 2: *Retelling*. Immediately following the Picture Walk, the book was taken from the child, and he or she was asked to retell as much of the story as possible. After the retelling, the child was asked if he or she could remember anything else about the story. Children's retellings were transcribed, and the information was categorized according to six main narrative elements: setting, characters, goal/initiating event, problem/episodes, solution, and resolution/ending. These elements were based on the story grammar framework (Mandler & Johnson, 1977). One point was awarded for phrases indicating the presence of each story element. Retelling scores ranged from 0 to 6, with 0 points demonstrating that the child did not recall any of the story elements and 6 signifying that the child's retelling included all of the elements.

Part 3: *Prompted comprehension*. The third part of the NC task is designed to assess the child's level of narrative comprehension, which is defined as the construction of meaning from pictures by integrating information across pages to create coherent and connected understandings. Following the retelling, the child was told that the experimenter and child would go through the story together a second time while the experimenter asked questions about the pictures. The experimenter guided the page turning and elicited reactions during this second viewing of the book by pointing out pictures and asking a series of 10 comprehension questions. Five of these questions were about explicit information and required the identification of characters, setting, initiating event, problem, and outcome resolution. Discussion of the latter three elements was followed by "why do you think so?" questions in order to promote responses that demonstrated narrative comprehension.

The remaining five comprehension questions required children to make inferences from the pictures about the characters' feelings, dialogues, causal relations, predictions, and themes. The questions about implicit information were also followed by "why" probes in order to distinguish between children who could make "shallow" inferences and those who could connect the inferences to other story events. (See Appendix B for prompted comprehension questions.)

Regardless of their responses to each question, children were always provided with one prompt (i.e., "Is there any other reason why the characters were feeling that way?"). Both the uniform prompts and the "why" questions provided an impetus for narrative

thinking and more elaborated responses to the book. This procedure ensured that all children were given opportunities to respond to the meaning of the story.

Scoring children's responses

A range of responses for each of the 10 prompted comprehension questions emerged from the pilot data, so we designed a scoring rubric that reflected story coherence. The rubric awarded more points for integration of information across pictures than for the description of a single picture in isolation. Thus, higher levels of narrative understanding were demonstrated by integration of information across pictures. This criterion of text cohesion across pages was applied to all 10 questions in the present study by using a 0–1–2-point rubric, which depended on what narrative comprehension versus focusing on single pages meant for each item (see Appendix C). In general, 0 points indicated no answer, irrelevant, wrong, or inappropriate answers; 1 point represented appropriate answers derived from single pictures; and 2 points were awarded when information from multiple pictures was used to create a coherent explanation that was consistent and connected with the child's unfolding narrative. Some new responses differed from the initial list of responses that was created during the pilot work, but it was always possible to use the three-level text cohesion criterion in order to categorize unique responses that children produced.

The 10 prompted comprehension questions yielded three composite scores: explicit comprehension, implicit comprehension, and total prompted comprehension. The explicit and implicit comprehension subscores ranged from 0 to 10 points. The total comprehension score included all 10 questions and ranged from 0 to 20 points. Higher scores on all three scales represent higher levels of narrative comprehension.

Interrater reliability

In order to ensure that the prompted comprehension rubrics yielded consistent scores across raters, 30% of the sample was randomly selected for an interrater reliability check. There was an equal representation of children by grade and reader or prereader status in this subsample. Two research assistants were trained to use the rubrics and scored children's responses to comprehension questions. Scores were checked for agreement by individual questions and tabulated by item and total percent agreement. Interrater reliability was above 90% agreement for every item with a mean of 97% agreement across all items.

Interrater reliability was also checked for the retelling measure. An independent rater was trained in the use of the rubric, and 30% of the retellings were randomly selected for scoring by the rater. Scores were checked for agreement by total retelling points, and the percent match was calculated. Retelling reliability was above 90%.

In order to evaluate the reliability of the Picture Walk, we selected an independent sample of 30 kindergarten, first-grade, and second-grade students (10 at each grade level) and administered just the Picture Walk section of the NC task. Two research assistants were trained on the Picture Walk administration protocol and observation scheme, and they simultaneously observed each of the 30 students during the Picture Walk and completed their own observation scheme and rated each child on the five behavior categories. Interrater agreement was calculated with the use of individual behavior and total scores for the 30 children. Percent agreement for book-handling skills, engagement, picture comments, storytelling comments, and comprehension strategies was 100%, 93%, 97%, 90%, and 97%, respectively. The overall percent agreement for the total Picture Walk scores was 95%.

The Qualitative Reading Inventory-II (QRI-II)

The QRI-II (Leslie & Caldwell, 1995) was used to assess five distinct reading skills: (a) reading of isolated words in graded word lists, (b) oral reading accuracy with miscue analysis, (c) retelling of the passage, (d) answering of comprehension questions, and (e) total reading time. The variables of interest were the word-list scores as well as the retelling and comprehension measures. Word lists were used to identify children as prereaders or conventional readers as well as to select appropriate-level passages and the retelling and comprehension measures that were used to examine the concurrent validity of the NC task for conventional readers. The passages were used to obtain the comprehension data, but oral reading measures were not used in analyses and so, therefore, are not described.

Graded word lists. The graded lists of 20 words in the QRI-II assessed children's abilities to read isolated words. Children were first given a word list at their current grade level and then a second list at the next grade level higher or lower—depending on their performance on the first word list. For each list, three scores were obtained: percentage of words identified automatically (less than one second), per-

centage of words identified with analysis (greater than one second), and total percentage of words correctly identified. Participants received two word lists, although some students received more than two word lists if their reading level was far enough away from the initial grade-level list that more lists were required to arrive at the child's actual level. Children who could not read the preprimer list did not receive a second, more difficult list.

Retelling and comprehension. Children orally read two of the graded reading passages. The first passage given to each child was selected on the basis of his or her score (90–100% correct) on the equivalent grade-level word list. The second passage was generally one grade level higher unless the student read the first passage below the frustration level (under 90% correct on the word list), in which case the child received the next lower level passage. Those who struggled to read the lowest level passage received no others unless they could complete the most elementary story (the preprimer level). After reading the selected passage, students were asked to retell as much of the story as possible and were given one prompt for any additional information they could remember at the end of the retelling. Retellings were initially scored according to the propositions supplied in the QRI-II manual so that students received a "propositions recalled" score indicating the number of propositions that the child could remember from the story. In addition, an alternative scoring system (key ideas) was created. It was based on narrative story structure, and children could receive 0–6 points depending on whether their retelling included information about six story elements: setting, characters, initiating event, problem, solution, and resolution/ending. Comprehension was assessed by children's answers to questions, and they were scored according to the QRI-II manual. There were five to eight questions per passage, and each set of questions included both inferential and literal questions.

Michigan Literacy Progress Profile (MLPP)

The MLPP (1998) was developed by the Michigan Department of Education to assess multiple features of children's early literacy. It includes assessments of five milestone skills: oral reading fluency, reading comprehension, writing, oral language, and attitudes and self-perceptions; and six enabling skills: concepts about print (CAP), letter-sound identification, phonemic awareness (PA), decodable word lists, known words activity, and hearing and recording sounds (HRS). Students

were assessed on three tasks—CAP, PA, and HRS—described in Appendix D.

Concepts about print. The CAP task assessed children's knowledge of the fundamental features of text as they examined printed text. The task contained 22 questions in six main categories: book concepts, reading concepts, directionality, concept of word, concept of letter, and punctuation marks. Questions resembled the following: "Show me the front of this book" or "Show me with your finger which way I go as I read this page." Students received 1 point for each correctly identified concept, which resulted in a CAP scale that ranged from 0 to 22 points. Higher points indicated more knowledge about printed text.

Phonemic awareness. The PA task assessed the child's understanding of the sound units of language. This task included three subsections: rhyming, phoneme blending, and phoneme segmentation. Children received three subscores ranging from 0 to 8 for each section, in addition to a PA total score, which was the sum of these three sections and ranged from 0 to 24 points. Higher scores indicated greater levels of phonemic awareness.

Hearing and recording sounds. The HRS task measured children's ability to hear individual phonemes and record them as letters. The task was a two-sentence story that children wrote after they heard the words. Students received 1 point for each correctly recorded sound; scores ranged from 0 to 39 because there were 39 distinct sounds in the two sentences.

Procedure

Children were assessed individually on the NC task. Some of the children were also assessed on either the QRI-II or on the MLPP tasks, depending on whether they had been identified as a "conventional reader" (for brevity, conventional readers will be referred to as readers) or a "prereader." After building rapport with each child the researcher gave the first QRI-II word list, with one word at a time exposed for the child to read. A higher level word list was given in the same fashion. No feedback was provided during the task, and all children were praised for their performance. If a child identified fewer than 12 of the 20 words (60%) on the preprimer word list, the student was recorded as a prereader and the remaining parts of the QRI-II were not administered. All other children were given two passages to read aloud that were based on the highest level of word list on which they identified at least 14 words. Children were told to read the stories as best

they could and that they would be asked questions about the story.

After the oral reading of each passage, retellings were initiated, and comprehension questions were asked. The readings, retellings, and comprehension questions were all audiorecorded for later scoring. If students who were given the preprimer passage could not decode the first few lines, the QRI-II was terminated and the child was recorded as a prereader. Hence, participants were identified as prereaders based on low performance on either the preprimer word list or the preprimer passage. The QRI-II was used as the criterion for identifying children as prereaders versus readers. According to the QRI-II manual, a child who identifies between 80 and 90% of the words from any given word list would read the corresponding grade-level passage at the "frustrational" level, making the passage too difficult and suggesting that a lower level passage be given to the child. We used the score of 12 on the preprimer word list (only 60% correct on the lowest possible level) as the cutoff for identification of readers and prereaders in order to apply a conservative criterion for categorizing children based on their reading abilities. The distinction between children who could read conventionally and those who could not was made for the purposes of our data analyses based on our methods and operational definitions, but we recognize that children's literacy development is continuous and lifelong. Administering the complete QRI-II took approximately 15 to 20 minutes per child.

After completion of the QRI-II, all students were given the NC task, using the *Robot-Bot-Bot* picture book (Krahn, 1979). Administering the three parts of the task required approximately 10 to 15 minutes per child. Two children did not respond to any of the Part III comprehension questions and were dropped from the analyses. After completion of the NC task, the three MLPP activities were administered to all children who had been identified as prereaders. The MLPP assessments required approximately 15 minutes per child. The tasks were often administered in separate sessions, depending on the attention of the child and the schedule of classroom activities.

Analytical techniques

Means and standard deviations were first considered for each of the five NC dependent variables in order to examine the distributions as well as the trends by grade. Correlations among NC variables were also examined in order to explore their patterns of relations. After these initial descriptive analyses,

each of the three NC phases was examined separately in order to examine performance patterns and the developmental sensitivity of each of the measures. For Picture Walk and prompted comprehension, the items were first examined individually in order to understand the specific breakdown of points for each item. Next, analyses of variance (ANOVA) by grade were conducted for Picture Walk, retelling, and prompted comprehension in order to test for main effects. Because preliminary analyses involving gender revealed no main effects or interactions with grade for any of the NC outcomes in all three studies, all subsequent analyses were collapsed across this variable. Reading ability differences in NC outcomes were also examined. However, reading ability effects could not be tested across grades because Grade × Reading Ability interactions could not be examined due to the near empty cells for kindergarten readers ($n = 1$) and second-grade prereaders ($n = 4$). To avoid confounding grade and reading ability, differences between prereaders and readers were examined only among the first graders. Cell sizes were insufficient to test for ethnicity differences, but we believe that the range of ethnicities represented should make the findings generalizable to other ethnically diverse populations. Additionally, intercorrelations were examined among NC outcomes and the MLPP measures for prereaders as well as among NC outcomes and QRI-II retelling and comprehension measures for readers.

Results and discussion

Each of the five NC task outcome variables was normally distributed with no ceiling or floor effects. Overall, there was regular improvement with age and reading ability on measures of retelling and prompted comprehension but few differences on measures of Picture Walk. As indicated in Table 2, there were significant positive correlations among the variables. Retelling and prompted comprehension scores were

TABLE 2
CORRELATION MATRIX FOR NC MEASURES IN STUDY 1

Variables	1	2	3	4	5
1. Picture Walk	—	.19*	.17*	.21**	.11
2. Retelling		—	.54**	.53**	.44**
3. Narrative comprehension total			—	.89**	.89**
4. Comprehension explicit subscore				—	.62**
5. Comprehension implicit subscore					—

Note. *$p < .05$. **$p < .01$.

TABLE 3
PERCENTAGE OF CHILDREN RECEIVING SCORES OF 0, 1, OR 2 FOR EACH PICTURE WALK BEHAVIOR

Dependent variable	0 points	1 point	2 points
Book-handling skills	0.6	9.5	89.9
Engagement	0.6	21.5	77.8
Picture comments	20.3	7.0	72.8
Storytelling comments	19.0	21.5	59.5
Comprehension strategies	54.4	29.7	15.8

more highly correlated with each other than either of them was with Picture Walk. The following sections describe children's performance on the three parts of the task.

Picture Walk results

Children received scores ranging from 1 to 10. Higher scores indicated more sophisticated interactions while independently viewing the picture book. Table 3 shows the percent of children scoring 0, 1, and 2 points on each of the Picture Walk behaviors. The percentage of children who received 2 points for each behavior category declined as the behaviors became more complex. Approximately 90% of the children scored 2 points for book-handling skills, whereas only 16% received 2 points for comprehension strategies. Picture Walk scores ranged from 2 to 10, with a mean of 7.21 (SD = 2.18). Table 4 shows the scores for children as a function of grade level and reading ability. An ANOVA with the Picture Walk total scores revealed no significant effects due to grade. The ANOVA by reading ability among first graders revealed no significant differences between prereaders and readers.

Retelling results

Retelling of key idea scores ranged from 0 to 6, according to the number of major story elements included in the retelling. The overall mean retelling score was 3.47 (SD = 1.85). As indicated in Table 4, mean retelling scores increased significantly with grade level. The ANOVA with retelling scores yielded a significant grade effect, $F(2, 155) = 15.10$, $p <$ .001. A Scheffé post hoc test indicated that second graders had higher retelling scores than children in the other two grades. The ANOVA by reading ability revealed no significant effects among first-grade readers and prereaders.

Prompted comprehension results

Children received total prompted comprehension scores ranging from 0 to 20 that reflected the extent to which responses integrated story information across pictures and events. Table 5 shows the percent of children scoring 0, 1, and 2 points on each of the five explicit and implicit comprehension questions. It is evident that most explicit questions were easier to answer than implicit questions. The most difficult items concerned an appropriate setting, prediction, and theme, whereas easier items queried character, problem, and outcome resolution. The means for total prompted comprehension, explicit comprehension, and implicit comprehension were 13.26 (SD = 4.05), 6.98 (SD = 2.34), and 6.23 (SD = 2.29), respectively. Because the three comprehension dependent variables were not independent from one another, separate ANOVAs for each of the outcomes were performed. ANOVAs by grade with the three dependent variables showed that older children scored significantly higher than younger children: total comprehension, $F(2, 155) = 26.35$, $p <$ .001; explicit comprehension, $F(2, 155) = 27.10$, p < .001; implicit comprehension, $F(2, 155) = 15.37$, $p <$ .001. Scheffé post-hoc tests showed that there

TABLE 4
DEVELOPMENTAL CHANGES IN NC TASK MEASURES

NC task variables	Grade level means			Reader status means for first graders only	
	K ($n = 34$)	1 ($n = 61$)	2 ($n = 63$)	Prereader ($n = 36$)	Reader ($n = 25$)
Picture Walk	6.9 (2.0)	6.9 (2.3)	7.7 (2.1)	7.0 (1.9)	6.7 (2.8)
Retelling	2.4 (1.8)	3.2 (1.8)	4.3 (1.6)	3.0 (2.0)	3.5 (1.4)
Total comprehension	10.2 (4.3)	12.6 (3.8)	15.5 (2.7)	11.8 (3.6)	13.8 (3.7)
Explicit comprehension	5.1 (2.6)	6.8 (2.1)	8.2 (1.6)	6.6 (2.1)	7.1 (2.0)
Implicit comprehension	5.0 (2.5)	5.8 (2.3)	7.3 (1.7)	5.2 (2.1)	6.7 (2.3)

*Note. SD*s in parentheses.

TABLE 5
PERCENTAGE OF CHILDREN RECEIVING SCORES OF 0, 1, OR 2 FOR EACH PROMPTED COMPREHENSION QUESTION

Dependent variable	0 Points	1 Point	2 Points
Explicit			
Character	6.3	20.3	73.4
Setting	32.3	22.8	44.9
Initiating event	10.8	62.0	27.2
Problem	11.4	23.4	65.2
Outcome resolution	4.4	40.5	55.1
Implicit			
Feelings	13.9	44.9	41.1
Causal inference	12.0	15.2	72.8
Dialogue	8.9	56.3	34.8
Prediction	25.3	37.3	37.3
Theme	38.0	24.7	37.3

were significant differences in total prompted comprehension and explicit comprehension between successive grades. For implicit comprehension, significant differences emerged between second grade and first grade and between second grade and kindergarten but not between kindergarten and first grade. In an additional analysis comparing explicit and implicit subscores, a 3 × 2 (Grade × Explicit and Implicit Comprehension) repeated measures ANOVA showed that the explicit comprehension subscore was significantly higher than the implicit comprehension subscore, $F(1, 155) = 15.19, p < .001$. Although the mean differences between explicit and implicit scores were greater for first and second graders than for kindergartners, the ANOVA did not reveal any significant interactions.

Additionally, an ANOVA between first-grade prereaders and readers showed that first-grade readers scored significantly higher than first-grade prereaders on total prompted comprehension, $F(1, 59) = 4.71, p < .05$, and implicit comprehension, $F(1, 59) = 6.99, p < .01$. There were no significant differences between first-grade readers and first-grade prereaders on explicit comprehension. Older children and children who could read text were better at integrating pictures in order to answer explicit and implicit comprehension questions. The finding that first-grade readers outperformed first-grade prereaders on total prompted comprehension and implicit comprehension suggests that narrative comprehension skills reflect proficient use of cognitive abilities and are not just consequences of age or grade.

Relations with other reading variables for prereaders

The intercorrelations among variables for prereaders were examined between NC task outcomes and MLPP measures in order to determine if enabling skills for reading are related to developing narrative competence. The MLPP tasks were not significantly correlated with Picture Walk or retelling measures, but they did show strong relations with prompted comprehension. Prompted comprehension was significantly correlated with phoneme segmentation ($r = .35, p < .01$), PA ($r = .33, p < .01$), HRS ($r = .33, p < .01$), and CAP ($r = .44, p < .01$). In a hierarchical regression on prompted comprehension, age was entered in the first step and was a significant contributor to prompted comprehension, $b = .16, p < .01; R^2 = .11, p < .01$. At step 2, Picture Walk, retelling, CAP, and PA were entered, causing an additional R^2 change of .23 ($p < .001$). Retelling and CAP were significant predictors of prompted comprehension, $b = .62, p < .01; b = .40, p < .05$, respectively. The overall model accounted for 35% of the variance in prompted comprehension. Apparently, entering age into the regression equation first attenuates any relationship between phonemic awareness and prompted comprehension—perhaps because age and phonemic awareness are so highly correlated. The regression results suggest that narrative comprehension skills may be related to age and linguistic enabling skills but that the developing comprehension skills are not simply a byproduct of the age and linguistic skills nor entirely accounted for by them.

The finding that PA and CAP were significantly correlated only with prompted comprehension (not Picture Walk or retelling) suggests that these skills are all developing conjointly in this age range, although it does not imply a causal relationship among them. Considering these findings with the developmental trends on other measures of the NC task, it appears that the Picture Walk has less developmental sensitivity, implying that children from 5 to 8 years old look at picture books similarly. Additionally, the strong relation between retelling and prompted comprehension suggests that the two processes are strongly correlated skills that underlie narrative comprehension. Furthermore, as shown by the regression, phonemic awareness does not overlap the same skills as prompted comprehension and retelling, which leads to the tentative conclusion that narrative comprehension may be an independent cognitive contributor to early literacy development.

Relations between NC and other reading variables for readers

Intercorrelations among variables for readers were examined between NC task outcomes and QRI-II retelling and comprehension measures in order to examine whether the NC task relates to comprehension in conventional measures of comprehension in an informal reading inventory. Neither Picture Walk nor prompted comprehension correlated significantly with the QRI-II comprehension or retelling measures. Significant relations emerged between NC retelling and QRI-II comprehension ($r = .29$, $p < .01$) and between NC Retelling and QRI-II retelling propositions ($r = .31$, $p < .01$) and QRI-II retelling key ideas ($r = .27$, $p < .05$). The lack of relation between prompted comprehension and QRI-II comprehension may be due to the different ways that comprehension is assessed in the two tasks. Prompted comprehension in the NC task was based on identifying elements and explaining their narrative relations, whereas the comprehension measure in the QRI-II awarded points for less complex and less genre-specific thinking. However, the significant relation between NC task retelling and QRI-II retelling illustrates the similarity of the processes and validates the use of retelling in the NC task as good evidence about story understanding.

Reprise

The NC task appears to be a useful quantitative measure of young children's narrative comprehension. It assesses young children's thinking and comprehension of narrative sequences without reliance on or confounding due to decoding skills. The procedures are appropriate for 5- to 8-year-old children whether they can read or not. Hence, children's early and emerging comprehension can be evaluated even though they may not be able to decode text. The NC task has the positive properties of assessment instruments identified by Stallman and Pearson (1990). It assesses reading while children are engaged with an authentic book and is consistent with the types of interactions that children have with parents and teachers around picture books. The task provides multiple measures of successful interactions with books in a brief period. Additionally, the task may provide both consequential and curricular validity because it can be aligned with instructional practices and can have positive consequences for children if teachers emphasize similar narrative elements and relations in their instruction with books. The task

can also be used by teachers for children in remedial reading programs or in bilingual reading programs because it reinforces comprehension skills when decoding skills may be unable to support reading of text. For example, the NC task makes clear the importance of encouraging comprehension strategies and eliciting storytelling comments, and it outlines the story elements that young readers should be able to identify and connect to the entire narrative.

Study 2

If narrative competence is a general characteristic of children's thinking and reading, it should be evident in a variety of materials and stories. The results of Study 1 are consistent with the findings of van Kraayenoord and Paris (1996), but the studies used different picture books, procedures, and children from different countries. However, a test of the generalizability of the NC task is still needed with identical procedures and subjects. Because the NC task was created initially with the use of a specific book (*Robot-Bot-Bot*, Krahn, 1979), it is necessary to determine whether the protocol for task administration, the types of questions asked, and the scoring rubrics can be applied to other picture books in a manner comparable to the original NC task. Thus, Study 2 tested the generalizability and the reliability of the NC task with three picture books.

Method

Subjects

Study 2 participants included the prereaders from Study 1 (excluding three students who did not complete all three storybooks) and a new, randomly selected sample of readers in order to have a wider range of grades and reading abilities ($n = 91$). As shown in Table 1, the children in Study 2 were primarily kindergarten and first-grade students; their ages ranged from 61 to 98 months ($M = 77$, $SD = 9$). Seventy-seven percent of Study 2 participants were prereaders. There were approximately equal numbers of females and males, and the sample was ethnically diverse: 58.2% Caucasian, 18.7% African American, and 6.6% Asian American.

Task

Two additional picture books were selected to test the NC task: Mercer Mayer's *A Boy, a Dog, and a Frog* (1992, Dial) and Fernando Krahn's *Amanda and*

the Mysterious Carpet (1985, Houghton Mifflin). In the frog book, a boy leaves a pond after unsuccessfully trying to catch a frog, which then becomes saddened by the boy's departure. The carpet book tells the story of a girl who receives a magic carpet that flies out of control. Similar to the robot book, both the frog and the carpet books used black line drawings with no accompanying words and had clear story lines with an obvious sequence of events and narrative structure. Several pictures were omitted in order to create the adapted versions of these books. The remaining 24 pages of the frog book and 26 pages of the carpet book were photocopied and assembled into book format with spiral bindings and covers. The only words in the adapted book versions were the title and authors' names on the front cover. For reference purposes, the three versions of the task will be identified as NC-R (the robot book), NC-F (the frog book), and NC-C (the carpet book).

The same three-part format (Picture Walk, retelling, prompted comprehension) was applied to the new books. The Picture Walk did not need to be modified for Study 2 books because the original observation scheme and scoring system can be applied to children's examination of any book. Retelling was also scored according to the same six narrative elements (setting, characters, goal/initiating event, problem/episodes, solution, and resolution/ending) but was adapted to the information in the frog and carpet books. The order and spacing of questions for prompted comprehension were modified for NC-F and NC-C. The same five explicit and implicit comprehension questions were used, but the order had to be changed so that the questions occurred on pages appropriate to the stories.

The only question that had to be completely changed for the modified versions of the task was the final "theme" question, because it was created specifically for each book. As in NC-R, appropriate questions were followed by explanatory probes (e.g., "Why do you think so?"), and children's responses were always followed up with one prompt for additional information. The same scoring rubric was used for the prompted comprehension questions for NC-F and NC-C. Higher levels of narrative understanding are represented by the integration of information across pictures rather than a focus on describing a single picture in isolation.

Consistent with the interrater reliability checks performed on the prompted comprehension questions for NC-R, 30% of the sample with equal grade and reading ability distributions was randomly selected to score prompted comprehension questions for NC-F and NC-C. Two trained researchers scored

the 10 comprehension questions, and scores were checked for agreement by individual question type and for total percent agreement. For both the frog and carpet stories, percent agreement by item ranged from 89% to 100%, with a 94% average agreement across all items. The rubric, therefore, was useful across picture books, allowing raters to make reliable judgments about the degree of narrative thinking represented in children's responses to the prompted comprehension questions.

Procedure

Study 2 was conducted in the fall of 1998 simultaneous with Study 1. Participants in the Study 2 subsample received three picture books rather than just the single robot picture book. After children were given the QRI-II word lists (and reading passages if they were readers), children who had been selected for Study 2 received the NC-R, NC-F, and NC-C. In order to control for order effects, the order of picture book administration was randomly varied. As in Study 1, children first engaged in the Picture Walk, then gave a retelling, and finally were asked the series of prompted comprehension questions. For some children, only one or two books were given in a single sitting; examiners then returned later that day or the following day to administer the remaining books.

Analytical techniques

Means and standard deviations were first considered for each of the five NC dependent variables for all three NC task versions in order to compare the distributions as well as the trends by grade. Intratask correlations were then examined in order to compare the patterns of relations of the dependent variables within each task. Intertask (within-subject) correlations were also calculated to ascertain whether similar scores were obtained by the same children across picture books. After initial descriptive analyses, performance on each of the three NC task phases was examined separately for the three NC task versions to compare performance patterns and the developmental sensitivity of each of the measures. For Picture Walk and prompted comprehension, the items were first examined individually in order to understand whether the specific breakdown of points for each item was similar across task versions. Next, ANOVAs with grade were conducted for Picture Walk, retelling, and prompted comprehension in order to test whether main effects of grade were similar for each of the task versions. ANOVAs were also per-

TABLE 6
DEVELOPMENTAL CHANGES IN NC MEASURES FOR NC-R, NC-F, AND NC-C

NC task variables	Overall means	Means by grade		
		K (*n* = 31)	1 (*n* = 46)	2 (*n* = 14)
Picture Walk				
NC-R	7.14 (1.98)	6.84 (2.05)	7.17 (1.98)	7.71 (1.77)
NC-F	7.27 (2.20)	6.70 (2.45)	7.28 (2.17)	8.43 (1.09)
NC-C	7.08 (2.25)	6.94 (2.21)	6.89 (2.48)	8.00 (1.18)
Retelling				
NC-R	3.05 (1.89)	2.42 (1.88)	3.04 (1.81)	4.50 (1.40)
NC-F	3.00 (1.77)	2.13 (1.43)	3.17 (1.76)	4.29 (1.64)
NC-C	2.99 (1.86)	2.16 (1.66)	3.07 (1.74)	4.57 (1.70)
Total comprehension				
NC-R	12.18 (4.09)	10.35 (4.42)	12.59 (3.51)	14.86 (3.46)
NC-F	11.88 (3.49)	9.87 (3.09)	12.70 (3.39)	13.50 (2.82)
NC-C	11.35 (4.35)	9.06 (4.20)	12.00 (3.84)	14.29 (4.01)
Explicit comprehension				
NC-R	6.38 (2.47)	5.10 (2.68)	6.89 (2.07)	7.57 (2.10)
NC-F	6.51 (2.04)	5.63 (1.99)	6.80 (2.02)	7.43 (1.60)
NC-C	6.46 (2.45)	5.06 (2.45)	7.00 (2.10)	7.79 (2.23)
Implicit comprehension				
NC-R	5.71 (2.34)	5.03 (2.56)	5.70 (2.09)	7.29 (1.98)
NC-F	5.37 (1.99)	4.23 (1.83)	5.89 (1.91)	6.07 (1.54)
NC-C	4.89 (2.40)	4.00 (2.27)	5.00 (2.32)	6.50 (2.21)

*Note. SD*s in parentheses.

formed on first graders only to examine patterns across the three task versions according to reader versus prereader status.

Results and discussion

The five NC task measures for each of the three book versions revealed normal distributions with no ceiling or floor effects. As shown in Table 6, the means and standard deviations for each of the measures were very similar across the three versions. As shown in Tables 6 and 7, the three picture books also yielded similar trends by grade and reading ability on all five measures. Scores increased with grade level for retelling and prompted comprehension but there were very few age-related changes in Picture Walk behaviors. The slightly lower comprehension scores for NC-C may indicate greater narrative complexity in that book, but the patterns of means are similar across all three books.

Intratask correlations were then investigated in order to assess whether the NC task outcomes showed similar relations to one another for the three versions of the task. The measures related to one an-

other in expected directions. With the exception of the correlations between the NC-R Picture Walk and comprehension variables, significant positive relations existed between them all. Particularly important were the consistent, strong relations between retelling and prompted comprehension, shown in Table 8, which ranged from $r = .46$ to $r = .61$.

Intertask correlations by outcome were also examined to ascertain whether children obtained similar scores across picture books. Significant positive correlations emerged for each Picture Walk, retelling, and total prompted comprehension measure across NC task versions. For Picture Walk, $r = .62$, $p < .01$ between NC-R and NC-F; $r = .53$, $p < .01$ between NC-R and NC-C; and $r = .69$, $p < .01$ between NC-F and NC-C. For retelling, $r = .64$, $p < .01$ between NC-R and NC-F; $r = .70$, $p < .01$ between NC-R and NC-C; and $r = .61$, $p < .01$ between NC-F and NC-C. Total comprehension correlated $r = .63$, $p < .01$ for NC-R and NC-F; $r = .62$, $p < .01$ for NC-R and NC-C; and $r = .58$, $p < .01$ for NC-F and NC-C. These high within-subject correlations show that children were consistent on each dependent variable across all versions of the NC task. Similar strong re-

lationships were evident within subjects when correlations among tasks were computed using Spearman's rank order correlations.

Picture Walk

Picture Walk items were examined individually in order to compare the breakdown of points awarded for each NC task version. These detailed frequency analyses revealed similar patterns in Picture Walk behaviors for all three picture books. As the Picture Walk behavior increased in complexity, greater percentages of children received 0 points for the behavior. For all three books, approximately 1–2% of children received 0 points for book handling, whereas approximately 59–69% of children received 0 points for comprehension strategies. Conversely, fewer children were awarded 2 points for the more complex behaviors. Approximately 85–89% of children received 2 points for book-handling skills, and roughly 11–17% scored 2 points for comprehension strategies. Tables 6 and 7 show Picture Walk scores for children on all three NC task versions as a function of both grade level and reading ability. ANOVAs by grade level revealed no significant differences for either NC-R or NC-C. However, Picture Walk scores did significantly increase by grade for NC-F, $F(2, 87)$ = 3.10, $p < .05$. These significant changes with age for NC-F may be due to more subtle details in this book. Picture Walk scores for first-grade prereaders as compared to first-grade readers exhibited similar trends, as demonstrated by the ANOVAs, which revealed no significant differences based on reading status for all three task versions.

Retelling

Retelling scores were also similar across NC task versions. Children recalled approximately three major story elements from each of the three picture books. As shown in Table 6, children demonstrated consistent age-related differences across NC task versions. Kindergartners scored approximately 2 points for each picture book retelling, first graders received roughly 3 points for their retellings, and second graders recalled approximately four to five main story elements. An ANOVA demonstrated that these grade differences were significant for NC-R, $F(2, 88)$ = 6.59, $p < .01$; NC-F, $F(2, 87)$ = 8.79, $p < .001$; and NC-C, $F(2, 88)$ = 9.71, $p < .001$. As in Study 1, readers and prereaders scored similarly on all picture books, including approximately 3.0–3.5 main elements in their retellings regardless of NC task version. Mean retelling scores among first

TABLE 7
NC TASK PERFORMANCE BY BOOK AND READING ABILITY FOR FIRST GRADERS

NC task variables	Prereader mean ($n = 35$)	Reader mean ($n = 11$)
Picture Walk		
NC-R	7.09 (1.79)	7.45 (2.58)
NC-F	7.20 (2.17)	7.55 (2.25)
NC-C	6.91 (2.51)	6.82 (2.48)
Retelling		
NC-R	3.09 (1.98)	3.45 (1.63)
NC-F	3.09 (1.80)	3.45 (1.63)
NC-C	2.97 (1.79)	3.36 (1.63)
Total comprehension		
NC-R	11.91 (3.60)	14.73 (2.15)
NC-F	12.40 (3.31)	13.64 (361)
NC-C	11.06 (3.84)	15.00 (1.79)
Explicit comprehension		
NC-R	6.63 (2.12)	7.73 (1.74)
NC-F	6.60 (2.09)	7.45 (1.69)
NC-C	6.49 (2.11)	8.64 (.92)
Implicit comprehension		
NC-R	5.29 (2.09)	7.00 (1.48)
NC-F	5.80 (1.78)	6.18 (2.36)
NC-C	4.57 (2.36)	6.36 (1.63)

Note. SDs in parentheses.

graders were not significantly different by reading ability for any of the NC task versions.

Prompted comprehension

The 10 prompted comprehension items were examined individually in order to compare the breakdown of points awarded for each NC task version. These frequency analyses revealed similar patterns in narrative comprehension levels for all three picture books. For all picture books, the highest percentage of students received 0 comprehension points for the setting, prediction, and theme questions. For NC-C only, many students (45.1%) could not infer characters' feelings. Furthermore, the majority of students received 1 point for initiating event and dialogue, and one quarter to one half of the children provided 1-point answers for most of the other comprehension questions. Patterns were also similar across picture books for receiving 2 points. The highest percentages of children scored 2 points for character, causal inference, problem, and outcome resolution questions, whereas fewer children received 2 points on the initiating event, prediction, theme, dialogue, and feeling questions. These scores indicate that many children

TABLE 8
INTRATASK CORRELATION MATRIX FOR NC-R, NC-F, AND NC-C

	Picture Walk	Retelling	Total comprehension	Explicit comprehension	Implicit comprehension
Picture walk	—				
A. NC-R		.23*	.20	.19	.18
B. NC-F		.43**	.43**	.38**	.36**
C. NC-C		.36**	.37**	.33**	.34**
Retelling		—			
D. NC-R			.47**	.49**	.33**
A. NC-F			.46**	.44**	.35**
B. NC-C			.61**	.55**	.54**
Total comprehension			—		
C. NC-R				.88**	.87**
A. NC-F				.87**	.86**
B. NC-C				.90**	.89**
Explicit comprehension				—	
C. NC-R					.56**
A. NC-F					.51**
B. NC-C					.61**

Note. *$p < .05$. **$p < .01$.

focused on events on a single page rather than the narrative level of meaning.

Examination of the prompted comprehension, explicit comprehension, and implicit comprehension distributions by grade level indicated similar trends across picture books (see Table 6). Significant age-related increases emerged for these three measures on each picture book. Older children received significantly more points than younger children on total prompted comprehension for all three NC task versions: NC-R, $F(2, 88) = 7.18$, $p < .001$; NC-F, $F(2, 87) = 9.17$, $p < .001$; NC-C, $F(2, 88) = 9.47$, $p < .001$. Grade-level differences emerged on explicit comprehension for NC-R, $F(2, 88) = 7.18$, $p < .001$; NC-F, $F(2, 87) = 5.10$, $p < .01$; and NC-C, $F(2, 88) = 9.47$, $p < .001$. Similar trends were revealed for implicit comprehension: NC-R, $F(2, 88) = 4.86$, $p < .01$; NC-F, $F(2, 88) = 8.64$, $p < .001$; and NC-C, $F(2, 88) = 5.88$, $p < .01$.

Means by reading ability among first graders were also comparable across picture books. Prereaders received approximately the same number of points on the three comprehension measures for all three picture books, and readers also scored approximately the same number of points on the three measures across picture books (see Table 7). ANOVAs by reading ability revealed significant effects for NC-R and NC-C but not for NC-F. Readers scored significantly more points than prereaders on total prompted comprehension for NC-R, $F(1, 44) = 5.99$, $p < .05$, and NC-C, $F(1, 44) = 10.72$, $p < .01$. For implicit com-

prehension, readers performed better than did prereaders for NC-R, $F(1, 44) = 6.32$, $p < .05$, and NC-C, $F(1, 44) = 5.50$, $p < .05$. Readers performed better than prereaders on explicit comprehension only for NC-C, $F(1, 44) = 10.69$, $p < .01$. Thus, there is some evidence that first-grade readers have better comprehension skills on the NC tasks. However, the effects were found on just one of the two new books, and the small cell size of first-grade readers ($n = 11$) means that the difference by reader status must be viewed with caution.

Study 2 demonstrates that the NC task yields remarkably consistent results across three different picture books. The similarity of developmental trends across books indicates that the NC task is sensitive to progressive increases in children's abilities to make inferences and connections among pictures and to construct coherent narrative relations from picture books. Similarity of performance across books shows that examiners can administer the task with different materials and score children's performance in a reliable manner. Thus, the generalizability and robustness of the NC task across picture books are supported.

Study 3

Studies 1 and 2 showed that the NC task is a reliable measure of children's understanding of narrative picture books with sensitivity to differences in

age and reading ability. These studies also demonstrated that the task correlates with other measures of early reading and that both the procedures of administration and children's performance are generalizable across multiple picture books. Study 3 had several purposes. First, we refined the administration procedures of the NC task and improved the manner of scoring children's retellings. Second, we examined the sensitivity of the NC task by analyzing within-subject growth over a one-year period according to children's age and reading ability. Because this study included some children who received the NC task and some who did not, it was also possible to investigate whether scores were inflated due to practice effects. Third, we tested the generalizability of the task with a new book and new subjects. Fourth, Study 3 tested the predictive and concurrent validity of the NC task with standardized and informal measures of reading.

Method

Subjects

Participants included 54 children who were kindergartners and first graders in year 1 (Studies 1 and 2) and 87 new children who were randomly selected for involvement in year 2 of data collection. As shown in Table 1, approximately equal numbers of kindergarten, first-, and second-grade children participated in Study 3; kindergartners were primarily prereaders, first graders were primarily readers, and all second graders were readers. Approximately equal numbers of females (n = 73) and males (n = 68) participated in Study 3, and the sample was ethnically diverse: 48% Caucasian, 24% African American, 16% Asian American, and 12% other or multiracial.

Task

Two picture books were used in Study 3, the identical robot book from Studies 1 and 2 and one new book, Mercer Mayer's *One Frog Too Many* (1992, Dial). This story is about a boy's pet frog that becomes jealous and attempts to get rid of a new baby frog that the boy has recently received as a gift. Similar to the books used in previous studies, this book had no accompanying words and a clear story line with an obvious sequence of events and narrative structure. Unlike the previous books, this book was in color, although the colors were muted and most of the page was white. An adapted version of the book was created by omitting several pictures

and by assembling the remaining 22 pages into book format with a spiral binding and cover. The front cover contained the only words in the adapted version of the book (title and author's name). For reference purposes, the two versions of the NC task used in this study will be identified as NC-R2 (to indicate robot book at Year 2) and NC-BF (to indicate baby frog book).

The three-part format was applied to NC-BF. No modifications were made to the Picture Walk observation scheme and scoring system. Retelling was scored according to the same six narrative elements (setting, characters, goal/initiating event, problem/episodes, solution, resolution/ending). In addition to the 0–6-point scale indicating the number of elements present, each retelling element was also coded for the amount of language and appropriate sequence expressed by each child. The following two variables were used to score the retellings: retelling word count (to account for the length of children's retellings) and retelling sequence (to describe the amount of narrative elements children recalled using the appropriate sequencing of events). A 0–3-point scale was created that differentiated between children who included zero narrative events (0 points), only beginning information or completely incorrect ordering of narrative events (1 point), appropriately sequenced information from two parts of the story, such as beginning and middle or middle and ending (2 points), and information in the appropriate order from the beginning, middle, and end of the story (3 points). Character and setting information were coded as beginning, initiating event, problem, and goal information (scored as middle) and outcome resolution, solution, and ending information (scored as ending). These additional scoring rubrics were also applied to NC-R2 and were the only modifications made to the robot book. In order to differentiate the three retelling measures, they will be referred to as R-NE, R-S, and R-WC (narrative elements, retelling sequence, and retelling word count, respectively).

The same five explicit and implicit questions were used for the prompted comprehension section, but the order and spacing had to be changed so that the questions were appropriate to the baby frog story. The wording of the theme question was as follows,

> Now I want you to think about everything that you learned from reading this book. What advice would you give to the boy or the frogs? Try to think of something to say so that the same thing that happened in this story doesn't happen again.

As in the other versions, appropriate questions were followed with explanatory probes (e.g., "Why do you think so?"), and children's responses were always followed with one prompt for additional information. The 0–2-point rubric was used to score the prompted comprehension questions. More points indicated higher levels of integration of information across pictures rather than describing single pictures in isolation.

Procedure

Participants were individually assessed on the NC-BF one year after data collection for Studies 1 and 2 (Spring 2000). In order to examine growth and practice on the NC task, first and second graders (at year 2) were also given NC-R2, where 54 of them received NC-R at year 1 and 32 did not receive any NC task at year 1. Additionally, consistent with Study 1, a subset of the children were also assessed on either the QRI-II or on the MLPP tasks, depending on whether they had been identified as readers or prereaders. Aside from the administration of the NC-R2 to the older students, the procedure was the same as in Study 1. After children were given the QRI-II word lists (and passages if they were readers), kindergartners were then given NC-BF, and first and second graders were given NC-BF and NC-R2. The order of the picture book administration for those who received two NC tasks was randomly varied in order to control for order effects. As in the previous studies, children initially engaged in the Picture Walk, then provided the retelling, and finally were asked the series of 10 prompted comprehension questions as they viewed the book for the second time. After completion of the NC tasks, prereaders were then given the MLPP tasks. Teachers gave the Gates-McGinitie Reading Test (GMRT) to second graders (year 1 first graders), and these scores were obtained for study participants. Additionally, scores on the Iowa Test of Basic Skills (ITBS) were obtained for third graders who were in grade 2 in year 1 of the study.

Analytical techniques

Six main sets of analyses were performed to address the purposes of Study 3. First, total means and means by grade were examined for each of the seven NC task variables on both NC task versions in order to investigate whether scores confirmed the patterns observed in Studies 1 and 2. Items were explored individually and by grade for Picture Walk, retelling, and prompted comprehension to examine whether

developmental trends were evident for each item as well as whether the items with which children had greater difficulty corresponded to the more difficult items from Studies 1 and 2. Factor and reliability analyses were also performed for the Picture Walk items in order to explore the strength of the Picture Walk scale as a unified construct. ANOVAs were then performed on each NC task outcome for both task versions to test whether trends by grade confirmed those found in the first two studies. Reading ability effects were not considered in Study 3 due to the small cell sizes at each grade level (seven kindergarten readers, nine first-grade prereaders, and zero second-grade prereaders). Second, growth and practice effects from year 1 to year 2 were examined. Repeated measures ANOVAs were used to test for significant within-subject increases as well as for grade-by-time interactions on the NC task from year 1 to 2. Third, to test for practice effects, a 2×2 (Practice $\times$ Grade) ANOVA was performed to examine whether the means were significantly higher for children who had previously received the NC task. Fourth, intertask correlations were examined to explore whether students performed similarly across picture books, which would further confirm the generalizability of the task. The fifth and sixth sets of analyses involved examination of correlations between the NC task and standardized and informal reading measures for evidence of the concurrent and predictive validity of the task.

Results and discussion

Each of the seven NC dependent variables for both NC-BF and NC-R2 had normal distributions with no ceiling or floor effects. As indicated in Table 9, significant positive correlations between variables were observed, and the retelling and prompted comprehension scores correlated more strongly with each other than with Picture Walk. The direction and relative strength of these relations confirm the same patterns found in Studies 1 and 2. The two additional retelling measures, R-WC and R-S, were strongly correlated with the number of narrative elements included in the retelling as well as with the prompted comprehension measures. R-WC, R-S, and R-NE were correlated more strongly with one another than with the prompted comprehension measures, as might be predicted from the similarities among the numbers of words, sequences, and narrative elements recalled.

TABLE 9
CORRELATION MATRIX FOR NC-BF AND NC-R2 MEASURES

Variables	Picture Walk	R-NE	R-S	R-WC	Total comprehension	Explicit comprehension	Implicit comprehension
Picture Walk							
NC-BF	—	.19*	.18*	.10	.13	.15	.08
NC-R2	—	.27*	.14	.25*	.25*	.24*	.19
R-NE							
NC-BF		—	.86**	.68**	.43**	.37**	.39**
NC-R2		—	.71**	.62**	.52**	.47**	.40**
R-S							
NC-BF			—	.65**	.54**	.46**	.48**
NC-R2			—	.62**	.50**	.44**	.40**
R-WC							
NC-BF				—	.46**	.43**	.37**
NC-R2				—	.44**	.45**	.30**
Total comprehension							
NC-BF					—	.88**	.86**
NC-R2					—	.82**	.85**
Explicit comprehension							
NC-BF						—	.53**
NC-R2						—	.40**

Note. *$p < .05$. **$p < .01$.

Confirmation of scores and patterns

The following sections present evidence from Study 3 that confirms the patterns of findings observed in Studies 1 and 2.

Picture Walk. The Picture Walk measure is an aggregated scale of different behaviors that does not represent a *theoretical* construct. For that reason, factor and reliability analyses were performed on the Picture Walk data for all five books (from all three studies) in order to examine the strength of the Picture Walk variable as a unified construct. Factor analyses on the Picture Walk items on all five picture books showed that for four of the five books, only one component was extracted when using varimax rotations and extracting all factors with eigenvalues greater than one (factor loadings on the four books ranging from .79–.92 for engagement, picture comments, and storytelling comments and ranging from .25–.59 for book-handling skills and comprehension strategies). Two components were extracted for the NC-R, with book handling being extracted by itself and engagement, picture comments, storytelling comments, and comprehension strategies extracted as the second component (factor loadings were .81, .85, .89, .53, respectively). Reliability analyses for the Picture Walk data from all five books revealed

the following high alphas: .69 for NC-R, .72 for NC-F, .72 for NC-C, .79 for NC-BF, and .75 for NC-R2. Although book-handling and comprehension strategies are more loosely related to the other items, the scales seem to hold together fairly well.

Individual examination of means and standard deviations for each Picture Walk item confirmed the patterns that were observed in Studies 1 and 2. As shown in Table 10, the means by grade and overall means decreased as the Picture Walk behavior increased in complexity for both NC-BF and NC-R2. For NC-BF, the mean for book-handling skills (the least complex behavior) was the highest for all grades, and the mean for comprehension strategies (the most complex behavior) was the lowest across all grades. The only location of deviation from this pattern of decreasing means with increasing complexity in behavior occurred with engagement and picture comments for kindergartners, but those discrepancies appear minor. For NC-BF, the pattern was confirmed for all five behaviors at all grade levels, with the mean for book-handling skills as high as 2.0 and for comprehension strategies as low as .5. In addition to these patterns in mean scores, the standard deviations for both NC versions increased across grades as the behaviors became more complex.

TABLE 10
PICTURE WALK MEANS ON NC-BF AND NC-R2

Dependent variable	Overall means	Means by grade		
		Kindergarten	Grade 1	Grade 2
Book-handling skills				
NC-BF	1.9 (.4)	1.9 (.4)	1.8 (.5)	1.9 (.3)
NC-R2	1.9 (.3)	—	1.8 (.4)	2.0 (.2)
Engagement				
NC-BF	1.7 (.5)	1.6 (.6)	1.7 (.5)	1.9 (.3)
NC-R2	1.8 (.5)	—	1.7 (.5)	1.8 (.4)
Picture comments				
NC-BF	1.6 (.8)	1.7 (.6)	1.3 (1.0)	1.8 (.6)
NC-R2	1.6 (.8)	—	1.4 (.9)	1.7 (.7)
Storytelling comments				
NC-BF	1.3 (.9)	1.3 (.8)	1.1 (1.0)	1.7 (.7)
NC-R2	1.4 (.9)	—	1.2 (.9)	1.7 (.7)
Comprehension strategies				
NC-BF	.8 (.8)	1.0 (.9)	.5 (.7)	1.0 (.8)
NC-R2	.5 (.8)	—	.6 (.8)	.6 (.8)

*Note. SD*s in parentheses.

Composite Picture Walk scores were very similar to those observed in the first two studies, with M = 7.3 (*SD* = 2.4) for NC-BF and M = 7.2 (*SD* = 2.4) for NC-R2. Table 11 shows scores for children on both tasks as a function of grade level. Whereas kindergartners' scores were not significantly different from those of first or second graders on the Picture Walk, the ANOVAs did reveal significant differences between first graders and second graders for both NC-BF, $F (2, 136)$ = 6.9, $p < .001$, and NC-R2, $F (1, 83)$ = 4.8, $p < .05$. These are significant but small differences and reflect the same trends evident in Studies 1 and 2.

Retelling results. Because each narrative element was specifically coded for its presence or absence in the child's retelling, it was possible to look more closely than in Studies 1 and 2 at the patterns that emerged for the individual elements. Table 12 shows that older students were more likely to include each of the six narrative elements in their retellings. Whereas four of the six narrative elements were recalled by less than 50% of kindergartners, most of the elements were recalled by well over 50% of older children. The elements that children were least and most likely to recall in their retellings were similar for both versions of the task; characters was the most difficult element, particularly for younger children, whereas setting was the most frequently included element across grades.

The overall R-NE score was M = 3.2 (*SD* = 2.0) of the total six elements possible for NC-BF and M = 4.3 (*SD* = 1.7) for NC-R2. ANOVA by grade showed that mean R-NE scores increased significantly with grade level for both NC-BF, $F (2, 137)$ = 9.3, $p < .001$, and NC-R2, $F (1, 84)$ = 4.9, $p < .05$. This was consistent with the trends in Studies 1 and 2. Similar grade increases were also observed for the two additional retelling measures: sequence and word count. On NC-BF, the score for retelling sequence was M = 2.0 (*SD* = .9), and scores increased significantly by grade, $F (2, 137)$ = 13.3, $p < .001$. Similar patterns emerged for NC-R2: M = 2.2 (*SD* = .9), and older students were significantly more likely to provide retellings with appropriate sequencing of events, $F (1, 84)$ = 5.9, $p < .05$. The retellings of older students were also significantly greater in length for NC-BF, $F (2, 136)$ = 10.0, $p < .001$, and for NC-R2, $F (1, 83)$ = 8.9, $p < .01$. Children on average provided retellings that were 91.7 (*SD* = 78.0) and 91.4 (*SD* = 56.9) words in length on NC-BF and NC-R2, respectively (see Table 11).

Prompted comprehension results. The means and standard deviations for each prompted comprehension question were first analyzed individually in order to examine whether the patterns confirmed those found in Studies 1 and 2. Table 13 shows that the easiest and most difficult questions for children in both NC-BF and NC-R2 were similar to those ob-

served in the earlier studies. For NC-BF, the easiest questions overall and in each grade were character, problem, dialogue, and feeling while the most difficult questions were theme, prediction, and initiating event. For NC-R2, the easiest questions overall and in each grade were character, problem, outcome resolution, and causal inference, and the most difficult questions were theme, prediction, initiating event, and feeling. The consistency in performance levels on narrative elements across books suggests that some elements of narrative may be more readily understood than other elements regardless of differences in story content and complexity.

In addition to confirming the differential difficulty of each of the prompted comprehension questions, individual examination of the means for each comprehension question showed increasing patterns of accuracy by grade for most of the 10 questions in both versions of the task. As shown in Table 13, first and second graders scored higher than kindergartners on all 10 questions, with the exception that kindergartners and first graders performed equally on the outcome resolution question in NC-BF. Second graders performed better than first graders on most questions, with the exception of one instance in each book where first graders barely outperformed second graders (dialogue question in NC-BF and character question in NC-R2). These patterns demonstrate that the developmental trends found in the previous studies were evident for most individual items.

For both books, overall means for the three prompted comprehension measures were consistent with the means found in Studies 1 and 2. NC-BF means were 14.0 (SD = 3.5), 7.1 (SD = 1.9), and 6.9 (SD = 2.0), and NC-R2 means were 15.7 (SD =

TABLE 11
DEVELOPMENTAL CHANGES IN NC MEASURES ON NC-BF AND NC-R2

NC task variables	Grade means		
	K	1	2
Picture Walk			
NC-BF	7.3 (2.3)	6.4 (2.6)	8.2 (2.0)
NC-R2	—	6.7 (2.5)	7.8 (2.3)
R-NE			
NC-BF	2.4 (2.0)	3.3 (1.8)	4.1 (1.8)
NC-R2	—	3.9 (1.8)	4.8 (1.6)
R-S			
NC-BF	1.5 (.9)	2.0 (.8)	2.4 (.7)
NC-R2	—	2.0 (.9)	2.5 (.8)
R-WC			
NC-BF	65.0 (65.9)	85.5 (57.8)	133.3 (95.9)
NC-R2	—	75.3 (44.3)	110.5 (64.3)
Total comprehension			
NC-BF	11.3 (3.2)	14.6 (2.6)	16.7 (1.8)
NC-R2	—	15.2 (3.0)	16.2 (2.1)
Explicit comprehension			
NC-BF	5.9 (2.0)	7.3 (1.5)	8.3 (1.2)
NC-R2	—	8.2 (1.7)	8.5 (1.3)
Implicit comprehension			
NC-BF	5.5 (2.0)	7.3 (1.6)	8.4 (1.1)
NC-R2	—	7.1 (1.8)	7.7 (1.4)

*Note. SD*s in parentheses.

2.7), 8.3 (SD = 1.5, and 7.3 (SD = 1.7) for total prompted comprehension, explicit comprehension, and implicit comprehension, respectively. Parallel to patterns observed in the first two studies, overall scores on the three prompted comprehension mea-

TABLE 12
PERCENTAGE OF STUDENTS WITH EACH NARRATIVE ELEMENT IN RETELLINGS

Grade	Setting	Character	Initiating Event	Problem	Resolution	Ending
Kindergarten						
NC-BF	62.7%	23.5%	54.9%	27.5%	37.3%	37.3%
NC-R2	—	—	—	—	—	—
First						
NC-BF	68.8%	35.4%	62.5%	56.3%	52.1%	58.3%
NC-R2	82.6%	58.7%	69.6%	58.7%	58.7%	65.2%
Second						
NC-BF	80.5%	61%	73.2%	68.3%	70.7%	56.1%
NC-R2	87.5%	70%	75%	77.5%	87.5%	75%
Total						
NC-BF	70%	38.6%	62.9%	49.3%	52.1%	50%
NC-R2	84.9%	64%	72%	67.4%	72%	70%

TABLE 13
MEANS FOR 0, 1, OR 2 SCALE ON PROMPTED COMPREHENSION QUESTIONS ON NC-BF AND NC-R2

Dependent variable	Overall means	Means by grade		
		Kindergarten	Grade 1	Grade 2
Explicit				
Character				
NC-BF	1.7 (.5)	1.4 (.6)	1.9 (.3)	1.9 (.4)
NC-R2	1.9 (.3)	—	2.0 (.2)	1.9 (.4)
Setting				
NC-BF	1.4 (.8)	.9 (.8)	1.6 (.7)	2.0 (.2)
NC-R2	1.6 (.6)	—	1.5 (.7)	1.8 (.5)
Initiating event				
NC-BF	1.2 (.5)	1.0 (.4)	1.3 (.6)	1.4 (.5)
NC-R2	1.3 (.6)	—	1.3 (.6)	1.3 (.6)
Problem				
NC-BF	1.5 (.6)	1.3 (.7)	1.4 (.6)	1.7 (.6)
NC-R2	1.7 (.6)	—	1.7 (.6)	1.8 (.6)
Outcome resolution				
NC-BF	1.3 (.8)	1.2 (.8)	1.2 (.8)	1.4 (.7)
NC-R2	1.7 (.5)	—	1.7 (.5)	1.8 (.4)
Implicit				
Causal inference				
NC-BF	1.3 (.6)	1.1 (.6)	1.4 (.6)	1.7 (.5)
NC-R2	1.8 (.6)	—	1.8 (.6)	1.8 (.5)
Dialogue				
NC-BF	1.7 (.6)	1.5 (.7)	1.9 (.4)	1.8 (.5)
NC-R2	1.5 (.6)	—	1.5 (.6)	1.6 (.6)
Feeling				
NC-BF	1.8 (.5)	1.6 (.7)	1.9 (.4)	1.9 (.4)
NC-R2	1.3 (.6)	—	1.3 (.5)	1.4 (.6)
Prediction				
NC-BF	1.1 (.8)	.6 (.7)	1.1 (.9)	1.6 (.7)
NC-R2	1.4 (.7)	—	1.3 (.8)	1.4 (.7)
Theme				
NC-BF	1.0 (.8)	.6 (.7)	1.0 (.7)	1.5 (.7)
NC-R2	1.3 (.8)	—	1.2 (.9)	1.4 (.7)

*Note. SD*s in parentheses.

sures for both books increased as a function of grade (see Table 13). ANOVAs revealed that, for NC-BF, older children scored significantly higher than younger children on total prompted comprehension, $F(2, 137) = 49.2$, $p < .001$; explicit comprehension, $F(2, 137) = 38.5$, $p < .001$; and implicit comprehension, $F(2, 137) = 38.5$, $p < .001$. Scheffé post-hoc tests showed that there were significant differences in all of the three comprehension measures between successive grades. For NC-R2, although there were increasing trends by grade on the three measures, these differences were not statistically significant, perhaps because this study had children from only two grade levels.

    The similarity in correlations, overall means, and developmental progressions of NC measures for NC-BF and NC-R2 confirms the patterns revealed by Studies 1 and 2 with new materials and additional children. Although frequency percentages by item and overall means for retelling and comprehension measures were slightly higher for NC-R2 than NC-BF, this can be attributed to the discrepancy in ages of children who received the two versions of the task, with NC-R2 not given to any kindergartners. Despite this difference in age ranges, similar patterns emerged, which lends support to the robustness of the NC task. Additionally, it is interesting to note that the Picture Walk means were not higher for NC-R2 than for NC-BF despite the age discrepancy, which again suggests that the Picture Walk is less developmentally sensitive than the retelling and comprehension measures.

TABLE 14

GROWTH ON NC MEASURES IN SUBSAMPLE OF CHILDREN RECEIVING NC TASKS AT YEAR 1 AND YEAR 2

Dependent variable	Year 2 grade	Year 1 mean	Year 2 mean	
		NC-R	NC-R2	NC-BF
Picture Walk	1	6.5 (2.4)	7.1 (2.3)	7.2 (2.3)
	2	7.5 (2.0)	7.8 (2.4)	8.3 (2.0)
	Total	7.2 (2.1)	7.5 (2.4)	7.8 (2.2)
R-NE	1	2.3 (1.6)	4.2 (1.7)	3.7 (1.9)
	2	3.4 (1.7)	4.8 (1.7)	4.2 (1.8)
	Total	2.9 (1.7)	4.6 (1.7)	4.0 (1.9)
Total comprehension	1	10.3 (4.7)	15.3 (3.1)	15.2 (3.0)
	2	13.3 (3.4)	16.3 (2.0)	16.7 (1.6)
	Total	12.1 (4.2)	15.9 (2.5)	16.1 (2.3)
Explicit comprehension	1	5.0 (2.6)	8.0 (1.7)	7.6 (1.7)
	2	7.0 (2.0)	8.5 (1.4)	8.4 (1.1)
	Total	6.2 (2.5)	8.3 (1.5)	8.1 (1.4)
Implicit comprehension	1	5.3 (2.6)	7.3 (1.9)	7.7 (1.6)
	2	6.2 (2.1)	7.8 (1.3)	8.3 (1.1)
	Total	5.9 (2.3)	7.6 (1.6)	8.1 (1.3)

*Note. SD*s in parentheses.

Growth and practice

In order to examine whether significant growth occurred in NC scores from Year 1 to Year 2, repeated measures ANOVAs (Time × Grade) were performed on a subset of the participants. The subset included first ($n = 22$) and second ($n = 30$) graders who received NC-R at year 1 and both NC-R2 and NC-BF at year 2. Table 14 shows scores on each of the five dependent variables at both time points according to grade level at year 2. (Other retelling measures were not given at year 1, so growth analyses are not possible.) As shown in Table 14, scores increased on all variables from year 1 to year 2 for both versions of the task, with Picture Walk scores increasing less than scores on the other variables. Although slightly increasing trends emerged on the Picture Walk, the differences were not significant for both NC-R2 and NC-BF, which again indicates that Picture Walk behaviors do not change much in this age range. The 2 × 2 (Time × Grade) ANOVA revealed significant growth on the remaining four variables for NC-R2: retelling, $F(1, 51) = 27.0$, $p < .001$; total comprehension, $F(1, 51) = 58.1$, $p < .001$; explicit comprehension, $F(1, 51) = 53.8$, $p < .001$;

and implicit comprehension, $F(1, 51) = 29.2$, $p < .001$. Significant growth was also observed on the four variables for NC-BF: retelling, $F(51) = 10.9$, $p < .01$; total comprehension, $F(1, 51) = 64.5$, $p < .001$; explicit comprehension, $F(1, 51) = 36.9$, $p < .001$; and implicit comprehension, $F(1, 51) = 50.0$, $p < .001$. For all four of these variables on both books, means for first graders increased more than second graders' means from year 1 to year 2. However, the ANOVA revealed only one significant Time × Grade interaction on explicit comprehension with NC-R2, $F(1, 51) = 6.3$, $p < .01$.

Practice effects were then examined in order to investigate whether the observed growth was confounded with previous exposure to the NC task in year 1. Post-hoc groups were created to compare first- and second-grade children who received the NC task at year 1 ("practice"; $n = 52$) with those who did not receive the NC task at year 1 ("no practice"; $n = 32$). The distribution by grade was uneven between the practice ($n = 22$ and $n = 30$ for grade 1 and grade 2, respectively) and no practice groups ($n = 24$ and $n = 8$ for grade 1 and grade 2, respectively). Therefore, a 2 × 2 (Practice × Grade) ANOVA was performed to

compare practice and no practice groups on the NC dependent variables for both task versions at year 2 so that potential confounds due to grade were considered. Though the means were somewhat higher for the practice group than the no practice group on most NC dependent variables, no significant main effects for practice or interactions with grade emerged between the two groups on any of the variables. ANOVAs revealed that significant grade effects accounted for any differences between practice and no practice groups. Therefore, it seems reasonable to conclude that the differences between the practice and no practice groups were not due to previous experience with the NC task in year 1 of the study.

The significant improvements in all retelling and comprehension measures for both books provide strong evidence of within-subject growth in narrative comprehension skills from year 1 to year 2. These clear patterns of growth provide additional evidence of the developmental sensitivity of the task as shown previously by the cross-sectional developmental analyses. Furthermore, the insignificant growth on Picture Walk for both books is consistent with the cross-sectional developmental analyses in the present study as well as those in Studies 1 and 2. This suggests that Picture Walk is less developmentally sensitive and that picture book viewing skills change little in this age range. Finally, it is worth noting that practice effects are minimal, especially when compared to longitudinal growth of NC skills.

NC task generalizability

The generalizability of the NC task was then assessed by comparing patterns of scores and relationships between the two Study 3 books with the subsample of first and second graders ($n = 83$) who received both tasks. Intertask correlations by outcome were examined in order to ascertain whether the same students received similar scores across picture books. As found in Study 2, significant positive correlations emerged between measures across NC task versions: Picture Walk, $r = .85$, $p < .01$; NE, $r = .60$, $p < .01$; total comprehension, $r = .51$, $p < .01$. Paired correlations between the two new retelling measures were also significant and very high: WC, $r = .77$, $p < .01$; order, $r = .62$, $p < .01$.

In addition to these significant within-subject intertask correlations, comparison of the patterns reported in the aforementioned sections of Study 3 further corroborates the generalizability of the NC task across materials and stories. Both tasks had very similar intratask correlations, which showed that children's performance on the NC task revealed similar relations between the dependent variables regardless of the book used. The detailed frequency analyses for Picture Walk, retelling, and prompted comprehension items also revealed very similar patterns of performance between books within narrative elements. Because the books differ in plot, style, graphics, and complexity, this consistency is remarkable. Additionally, the overall means for the two books were very similar on all measures, with only two of the means being significantly different when the same children were compared (R-NE and implicit comprehension). It is important to note that these comparisons between NC-R2 and NC-BF are between children in grades K–2 for the baby frog book and only grades 1 and 2 for the robot book. The patterns and scores are much more similar when only the subsample of children who received both is compared (thus equalizing the groups by dropping the youngest and poorest readers). The NC task also shows tremendous consistency in developmental sensitivity across books. Similar patterns emerged by grade for all dependent measures, and very similar patterns of growth were revealed from year 1 to year 2 on all of the measures. Furthermore, the findings from Study 3 confirm the results from Studies 1 and 2. This additional confirmation of similar NC performance between the two years with a new book also provides compelling evidence of the robustness of the NC task across children, time, materials, grades, and testers.

Concurrent validity

Relationships of NC-BF outcomes with both prereader (MLPP) and reader (QRI-II, GMRT) variables were then examined in order to confirm the concurrent validity of the NC task. The NC-BF was chosen over the NC-R2 because all children received the NC-BF and it was not confounded by practice. The pattern of intercorrelations among MLPP tasks and NC-BF dependent variables ($n = 54$) was similar to the pattern observed in Study 1. MLPP tasks were not significantly correlated with NC Picture Walk or the three retelling measures, but strong relationships between the MLPP tasks and prompted comprehension emerged. Prompted comprehension was significantly correlated with

TABLE 15
CORRELATIONS AMONG NC-BF VARIABLES AND MEASURES FOR READERS

Dependent variable	QRI-II propositions	QRI-II key ideas	QRI-II comprehension	GMRT vocabulary	GMRT comprehension
Picture Walk	.19	.08	.13	.00	.13
R-NE	.43**	.39**	.09	.20	.20
R-S	.45**	.36**	.12	.10	.04
R-WC	.54**	.51**	.18	.43*	.40*
Total comprehension	.30**	.33**	.26*	.39*	.53**
Explicit comprehension	.21	.25*	.15	.29	.43*
Implicit comprehension	.28**	.30**	.30**	.30	.38*

Note. *p <.05. **p <.01.

phoneme blending ($r = .34$, $p < .05$), rhyming ($r = .37$, $p < .01$), phonemic awareness total ($r = .42$, $p < .01$), hearing and recording sounds ($r = .55$, $p < .01$), and concepts about print ($r = .48$, $p < .01$). These relationships were stronger than those observed between MLPP and prompted comprehension in Study 1, particularly the relationships of prompted comprehension with CAP and HRS. The reasons for the strong correlations are not clear. The phonological skills and narrative skills could be developing concurrently in this age range or they might both be related to increasing reading skills. It seems unlikely that they contribute directly to each other.

Intercorrelations were then examined between NC-BF outcomes and QRI-II retelling and comprehension measures ($n = 87$) in order to investigate whether picture-based retelling and comprehension were correlated with retelling and comprehension performance as measured by a text-based task. As observed in Study 1, Picture Walk did not significantly correlate with QRI-II comprehension or retelling measures. As indicated by Table 15, the three NC retelling measures correlated significantly with the QRI-II retelling measures, which are more strongly correlated with QRI-II retelling propositions recalled than with QRI-II retelling key ideas. NC retelling measures did not correlate, however, with QRI-II comprehension questions. NC total comprehension and NC implicit comprehension were significantly correlated with QRI-II retelling and QRI-II comprehension, whereas NC explicit comprehension correlated significantly with QRI-II retelling key ideas but not with QRI-II retelling propositions or comprehension. The minor discrepancies in correlations with comprehension measures may reflect nuances among the small number of questions for each QRI-II passage. The overall pattern of results revealed many significant and positive

relations between comprehension on NC tasks and QRI-II passages.

Table 15 also shows the intercorrelations between NC-BF measures and GMRT vocabulary and comprehension scores. NC Picture Walk, R-NE, and R-S were not correlated significantly with GMRT vocabulary or comprehension subscores. Narrative comprehesion R-WC, total comprehension, explicit comprehension, and implicit comprehension were all significantly correlated with GMRT comprehension, and only R-WC and total comprehension were also significantly correlated with GMRT vocabulary. As indicated in Table 15, R-WC was more strongly correlated with GMRT vocabulary, and Total Comprehension was more strongly correlated with GMRT comprehension. These patterns of relations provide compelling evidence of the concurrent validity of the NC comprehension measure. It is particularly noteworthy that the strongest correlation that emerged among the 35 correlations in Table 15 was between NC total comprehension and GMRT comprehension. Furthermore, the two NC comprehension subscores were significantly correlated with GMRT comprehension and not with GMRT vocabulary. This suggests that comprehension processes are similar in the NC and GMRT assessments and that NC comprehension is not simply a reflection of children's vocabulary.

Predictive validity

In order to consider the predictive validity of the NC task, correlations were examined between the NC task at year 1 (NC-R) and both standardized (GMRT, ITBS) and informal (QRI-II) measures of reading comprehension at year 2. The ITBS was administered to 47 third graders at year 2 who had received NC-R in year 1. Significant correlations emerged between NC outcomes and ITBS subscores. NC measures (Picture Walk, retelling, total

comprehension, explicit comprehension, implicit comprehension) were more strongly correlated with the ITBS comprehension subscore (.30*, .46**, .52**, .54**, .35*, respectively) than with the vocabulary subscore (.30*, .23*, .41**, .42**, .28, respectively). The higher correlations of NC comprehension with ITBS comprehension than with ITBS vocabulary provide some evidence of the predictive validity of the NC task for children's subsequent comprehension skills. Correlations were also examined for the 44 second graders who received NC-R at year 1 and the GMRT at year 2; however, no significant relationships emerged between any of the variables. NC retelling, total comprehension, and explicit comprehension at year 1 were significantly related to QRI-II retelling propositions (r = .27, p <.05; r = .28, p < .05; r = .29, p < .01, respectively) and QRI-II retelling key ideas (r = .23, p < .05; r = .26, p < .05; r = .25, p < .05, respectively). NC Picture Walk was not significantly related to QRI-II outcomes, and none of the NC outcomes correlated significantly with QRI-II comprehension.

It would be useful to conduct a regression analysis with these data to predict comprehension scores with NC measures, but the sample sizes were simply too small. Also, children received either MLPP or QRI-II tasks, so both cannot be entered in the same analysis. However, there is some evidence that MLPP measures predict year 2 scores on reading assessments. For example, year 1 CAP and year 2 QRI-II comprehension correlated at r = .36; year 1 HRS and year 2 QRI-II key ideas correlated at r = .49; and year 1 MLPP blending and year 2 GMRT vocabulary correlated at r = .48. All these correlations were significant at p < .01, and the relations indicate that prereading skill assessments are related modestly to some reading assessments one year later. These relations, however, appear less compelling than the correlations between NC measures and later reading assessments—particularly later measures of comprehension.

In sum, Study 3 provided clear evidence of the robustness of the NC task. Study 3 confirmed the patterns that emerged in Studies 1 and 2 in year 1—showing additional evidence of the developmental sensitivity of the task by grade, the generalizability of the task across materials, and the concurrent validity of the task. Study 3 also provided a more thorough evaluation of children's retellings as elicited by pictures with the additional measures and the detailed analyses by individual story grammar element. Finally, the study demonstrated support for the predictive validity of the task as well as within-subject growth over time on NC measures.

Summary of findings

The three studies in this article provide consistent and positive evidence about the NC task as a measure of children's narrative understanding of picture books. All three studies indicate that the NC task is developmentally appropriate for 5- to 8-year-olds. The NC task provides children with authentic, enjoyable book experiences that are aligned with parent-child shared book interactions in the home. The materials as well as the administrative procedures resemble joint book-reading experiences that are familiar and motivating. For example, the picture book with a binding and cover is genuine because children can hold and manipulate the book as they normally do. Also, the task fosters guided page turning; responding to the pictures elicited by questions; talking about characters, settings, feelings, and themes; provision of character dialogue; and encouragement by adults for children to talk about connections among narrative events that are presented in picture books. We designed multiple measures of comprehension because they provide several different windows into children's comprehension of narrative, and each of them is appropriate for children in this age range. The Picture Walk allows children to become familiar with the book and story as well as with the assessment situation. Retelling is a common method of comprehension assessment that is like storytelling and familiar to children. Because the explicit and implicit comprehension questions are embedded in the viewing of the book, successful responses to questions minimize memory demands. The absence of floor or ceiling effects on all three of these measures further confirms their developmental appropriateness. The NC task is easy to use by researchers or teachers and can be given in less than 15 minutes, which is critical for individual assessment of children in this age range.

Second, the patterns of performance across grade levels demonstrate the developmental sensitivity of the measures. Retelling and prompted comprehension scores improved significantly with age, indicating that the NC task differentiates between children who can recall main narrative elements, identify critical explicit information, make inferences, and connect information across pages from children who have weaknesses with these narrative comprehension skills. The little evidence of change with age in Picture Walk behaviors suggests that young children interact with narrative storybooks similarly in K–2. Picture Walk performance may reflect how a beginning reader approaches and reacts to storybooks, which might have a greater link with previous experience, motivation, or

the specific context in which the task is given, but, nonetheless, it provides important descriptive information for all children in this age range. The Picture Walk is also important in the NC task for the child to become familiar with the entire book and story, so, even if it is not used as data, the Picture Walk cannot be deleted from the task.

Third, significant relations between the NC task and measures for readers and prereaders show some evidence of the concurrent validity of the task. The strong correlations between NC retelling and QRI-II retelling measures provide encouraging support for the validity of eliciting retellings from picture narratives. This finding suggests that the retelling, which is a popular measure of comprehension for young children, can be extended beyond decoding and listening tasks to picture formats. The correlations among NC and MLPP variables indicate that emergent reading skills might be linked, but the lack of relations between phonemic awareness scores and comprehension when entering age into a regression suggests the possible independence of narrative comprehension from phonological skills.

Study 2 provides compelling evidence of the generalizability of the NC task across multiple picture books that vary in content, plot complexity, graphic style, and length. Each of the NC measures and their corresponding administration procedures were applied to three different picture books. The protocol, observation scheme, and rubrics for the Picture Walk were transferable across the three books in this study without any modifications, and the identical procedures and target behaviors for this measure are applicable to many narrative picture books. Similarly, the identical method of eliciting a retelling following independent viewing of the picture storybook was used for all three of the books. The retellings for the three books could be scored with the same narrative element categories—the only necessary modifications were in scoring each of the narrative elements based on content specific to each of the stories. The same method of asking the embedded comprehension questions, as well as the types of questions asked, were also easily applied to all three books by changing the order and spacing of some of the questions so that they were aligned with the narrative structure of the story. The scoring rubrics and general rules for differentiating the quality of responses were also easily adapted for all three books.

In addition to the generalizability of the measures and procedures, the generalizability of the task was also evident in the scoring of NC task performance. The high percentage of agreement between raters across the three books showed that the scoring

rubrics are reliable regardless of differences in content and raters. The patterns of performance further confirmed the generalizability of the task. High intertask correlations between books for each of the measures showed that children scored similarly across books. Comparison of scores across the three books revealed strong similarity in overall means as well as very similar trends by age. Not only were the overall scores very similar, but the patterns on individual items for both Picture Walk and prompted comprehension also showed high consistency, with similar items emerging as easy or difficult across books. Overall, Study 2 shows that the NC task and measures can be used with multiple storybooks that present narrative information in pictures. The remarkable robustness of the NC task across picture books makes it a very accessible and adaptable assessment for both researchers and classroom teachers.

Study 3 offers more evidence of the reliability and validity of the NC task one year later with an additional book, allowing for the investigation of a new sample of kindergartners as well as longitudinal and predictive analyses that capitalized on two years of NC data from the same children. First, Study 3 demonstrated the reliability of the task by confirming the means and patterns that were found in Study 1. Overall means on the five measures for both books used in Study 3, in addition to the individual means by item, were very similar to the means that emerged in the previous studies. The easiest and hardest items in the NC tasks in Study 3 were very similar to the easiest and hardest items in the NC tasks in Studies 1 and 2. The patterns of performance also paralleled the patterns in the first two studies. Intratask correlations were positive and significant among the retelling and comprehension variables, with the Picture Walk being considerably less correlated with these variables. The developmental patterns by grade that were found in Study 1 were also evident for both Study 3 books, where significant increases were observed for retelling and prompted comprehension and fewer differences were found for Picture Walk.

Second, additional details about narrative were examined in Study 3, yielding a richer description of children's performance on the retelling measure. Item analyses of the specific narrative elements, rather than merely tabulating total scores, revealed consistency across the two books in the easiest and most difficult narrative elements. The addition of the sequence measure provided information about the proper ordering of events, which is an important aspect of retelling skill that was not accounted for by the narrative elements measure. This measure

showed good evidence of retelling performance, as demonstrated by its reasonable relations to other measures and its sensitivity to grade. The addition of the word count measure accounted for differences in the length of children's retellings, which was very related to all of the NC dependent variables except for Picture Walk. The stronger correlations of word count with the other retelling measures than with the comprehension measures suggest that retelling measures might be more aligned with language assessments, and comprehension measures more aligned with cognitive processing measures.

Third, examination of first and second graders who received the NC task over two years provided evidence of growth on the NC measures. The growth that emerged for both books was consistent with the developmental trends that emerged in the cross-sectional analyses, with significant growth occurring for retelling and prompted comprehension but not with Picture Walk. These analyses offer more support for the developmental sensitivity of retelling and comprehension and the lack of sensitivity of Picture Walk. Fourth, it was possible to explore practice effects due to viewing the same book in the prior year. Little evidence of practice effects was observed.

Fifth, positive and significant relations were found among the NC task and other measures of early reading. The NC task was significantly correlated with the following concurrent measures: MLPP phonological awareness tasks, QRI-II retelling, QRI-II comprehension, GMRT vocabulary, and GMRT comprehension. The NC task at year 1 also significantly predicted scores in year 2 on QRI-II retelling measures and ITBS comprehension and vocabulary. These strong concurrent and predictive relations, and in particular the stronger correlations of NC variables with comprehension as compared to vocabulary scores on the standardized reading tests, provide encouraging evidence of the validity of the NC task as a measure of comprehension for emergent readers.

Finally, Study 3 adds more evidence for the generalizability of the NC task across storybooks with a new book, new children, and a time lapse of one year. The measures and administrative procedures were easily applied to the new book, and consistent scoring was observed across raters in applying the rubrics. The high intertask correlations between the two books on all seven variables, including the two new ones, show that, similar to Study 2, children scored consistently on the dependent variables. Additionally, the similarity between books in the intratask correlations, overall means, means on individual items, and developmental trends and growth also supports the generalizability of the task. Study 3

findings corroborated the reliability, validity, and generalizability of the NC task.

Implications for research

We believe that narrative competence may be a general feature of children's thinking that is critical for early literacy and cognitive development. It is pervasive in children's language, television viewing, playing, reading, viewing of picture books, listening, and thinking (Bruner, 1986; Yussen & Ozcan, 1996), and it is supported by parents and teachers alike in their everyday practices. The three studies in this article provide compelling data to support recent claims that children's understanding of narrative is an important foundation for learning to read (Burns, Griffin, & Snow, 1999; Whitehurst & Lonigan, 1998). The data extend and confirm previous research with wordless picture books and show that narrative comprehension skills may reflect skills that are fundamentally important for reading development yet independent from language and decoding skills. Because most theories of early reading neglect to include narrative understanding among the prereading or emergent skills that precede conventional reading, more research is needed in order to calibrate the role of narrative understanding in these theories.

This research validates the narrative comprehension task as one method to conduct research on narrative understanding as a precursor to and correlate of beginning reading. The NC task yields reliable, quantifiable data through standard procedures that are generalizable across picture books. Researchers can use the uniform procedures of the NC task for comparisons with other reading processes in subsequent tests of developmental models of reading. Narrative competence can be expanded beyond the studies in this research in several ways. First, more tests of the long-term predictive validity of the NC task are needed. We plan to extend the present studies by examining how well narrative comprehension predicts children's comprehension on literacy tasks over greater time intervals. Because the NC task does not depend on decoding skills, nor does it require knowledge of the English language, research is also needed on the reliability and validity of these tasks with bilingual and non-English-speaking children.

Because the development of narrative comprehension depends so critically on children's experiences prior to formal schooling, it is important to extend the research on narrative comprehension to preschoolers in order to investigate the developmental appropriateness, validity, and diagnostic sensitivi-

ty of the task with 3- and 4-year-old children. If the NC task provides valid and reliable information about narrative thinking of preschool-age children, the task would be a particularly useful early diagnostic assessment.

Research on narrative thinking skills should be extended to include other contexts in which young children demonstrate narrative thinking. One of these ways is the media (e.g., cartoons, television shows, children's films, CD-ROM talking books, and other interactive media forms), where, similar to picture books, children must make sense of narrative information that is presented in a visual format. There are three important questions: (a) Can the NC task be applied to media formats to yield reliable and valid assessments of narrative comprehension? (b) Is comprehension of narrative in visual media related to comprehension of narrative storybooks? and (c) Is narrative comprehension of media related to emergent literacy skills and reading development? Preliminary evidence from studies of preschoolers' viewing skills suggests affirmative answers to all three questions (van den Broek, 2001).

Future research can assess children's storytelling abilities in addition to their comprehension. Narrative productions in children's language, play, and writing may provide additional indices of the development and quality of narrative thinking. We speculate that assessments of children's narrative thinking may show developmental advances in their narrative productions in both language (e.g., story telling) and print (e.g., story writing). We expect that even 2- to 4-year-old children can exhibit some aspects of narrative thinking with pictures, objects, and language. The idea that these experiences may influence literacy development is certainly a viable hypothesis and deserves further exploration.

Research is needed on the effectiveness of instructional interventions to teach narrative thinking. It is critical to understand whether children who receive narrative instruction improve their narrative comprehension skills. For example, does their knowledge of story grammar increase, do their sequencing and inferential skills improve, are they better able to integrate information, and do they exhibit more strategies for integrating meaning? Beyond these short-term effects, it is important to study whether narrative interventions facilitate learning to read and promote the development of reading comprehension. It is also important to determine the ages and abilities of children who benefit most from such interventions. This type of research on the instructional validity of the NC task is urgent because it may actually help children learn to read and un-

derstand better. Clearly, there are many ways to use the NC task in research. Overall, the NC task has the potential to redress the decoding and comprehension imbalance in early reading research, and it is a model of the kind of comprehension assessment tasks that are needed.

Implications for classroom practices

The NC task does not require decoding skills and can be adapted to many narrative books (e.g., texts that have an obvious sequence of events and main narrative elements), so teachers in primary grades can use the task with basal readers, oral stories, or visual media. We think that teachers can use the NC task interchangeably for assessment and instruction, which, in several ways, is a hallmark of an educationally useful and authentic task. First, K–3 teachers need comprehension assessment tools such as the NC task. The unbalanced attention in theory and research afforded to decoding as compared to comprehension skills extends to early literacy assessment practices in the classroom. Many early literacy assessments focus on reading skills that are decontextualized from the reading process (Stallman & Pearson, 1990). Assessments of enabling processes, such as decoding skills and phonological awareness, are frequently used exclusively in primary grades. Stallman and Pearson's review of the types of skills measured by early reading assessments revealed that the majority of the assessments focused on decoding and phonological skills, in contrast to the almost complete neglect of comprehension and language skills. When such skills are assessed, language, vocabulary, and print and book concepts usually receive the most attention. When comprehension is assessed, typically low-level skills are measured that involve recognition and recall, and they are usually at the word or sentence level. Typical comprehension measures do not assess the complex thinking required to actively construct meaning and understand narrative stories.

Beyond the paucity of available early comprehension assessments, there are few methods and opportunities to assess comprehension in primary grades. Teachers in K–2 grades who have students unable to decode words rarely measure story comprehension, either formally or informally. When it is assessed, teachers usually ask children a series of questions after listening to an orally presented story (Morrow, 1990), but such listening assessments are rarely tied to text difficulty, higher order thinking, or records of progress. We believe that the NC task exemplifies the types of complex, authentic tasks that

are needed to measure the development of narrative thinking skills.

Because the NC task has excellent diagnostic sensitivity, teachers may adapt it to informal classroom use in order to make their own diagnostic assessments of children's cognitive understanding of stories. Several strategies are possible. One way is to use the NC task to measure children's familiarity with books and their comfort with book-handling skills and conversations in the beginning of the year. A second method is to record scores for specific measures such as Picture Walk, retelling, and prompted comprehension, as described in Study 1, in order to record improvements in children's narrative thinking. A third method is to use the NC task as an early screening test of children's comprehension. A fourth method is to use the NC task as a specific diagnostic tool to assess children's abilities to infer events, verbalize characters' feelings, or summarize a sequence of events. We suggest collecting these kinds of information about children's performance by using multiple books throughout the year so that children gain practice thinking about narrative stories. This helps to align instruction and assessment practices with picture books.

The following decisions and steps may help teachers use the NC task effectively.

1. Choose a book with a clear goal-directed plot and pictures that convey the key ideas.

2. Decide what information to collect from children.

3. Allow children to read the book multiple times as behaviors are observed or recorded.

4. Use the observations to identify specific problems exhibited by children, such as poor book-handling skills, lack of inferences that connect meaning across pictures, or incomplete retellings.

5. Use the same or different books to reassess children during the year to determine if children demonstrate appropriate growth.

6. Share the information about children's emerging comprehension skills with children, parents, and teachers so that instruction can be connected to the observed strengths and weaknesses.

Many children begin school with poor narrative skills (Feagans, 1984), and they are likely to have difficulty in school in general—especially with literacy (Roth, 1986). Because beginning reading instruction in school usually has a heavy emphasis on decoding skills, teachers need to design complementary practices that foster narrative thinking skills. Teachers can focus lessons on the structural elements of narrative and how they are organized. They can

facilitate thinking about characters' internal states by encouraging children to ask themselves questions such as "How did that make the little boy feel?" or "What do you think the hungry dog is looking for?" Teachers can help children see that stories can be interpreted in multiple ways and that they must add their own interpretations in order to fully understand pictures or text. Direct instruction can be coupled with instruction that is embedded in genuine reading experiences, and teachers can create environments where children are engaging in conversations about the features of narrative and what they must do in order to read them successfully.

The NC task and its three-part format can serve as an effective instructional tool for teachers because it highlights the types of skills and thinking that should be promoted. For example, the Picture Walk highlights five categories of book-viewing skills to be promoted. Retelling underscores the main narrative elements to promote as well as the importance of sequencing events. Prompted comprehension emphasizes numerous narrative skills that teachers should foster (e.g., inferential skills including predictions, themes, and dialogue; knowledge about setting and characters). In addition, the elaborating prompts that the NC task uses to encourage children to integrate information can cue teachers to instruct children that they should ask themselves "why," "what next," and "how do you know" questions in order to comprehend stories thoroughly.

We think that a developmental framework can be applied to narrative comprehension instruction in classrooms where developmentally appropriate materials are adapted for the needs and skill levels of the classroom or a specific child. For example, preschool and kindergarten programs might use picture books, first graders and beginning readers could use storybooks with both pictures and text, and second graders might be instructed using storybooks with a mix of pictures and text before moving to text only. One of the greatest strengths of the NC task (and the instruction that it encourages) is its flexibility—it can be modified easily to match the developmental and personal needs of any child. In addition to using developmentally appropriate materials, direct instruction of narrative comprehension skills should be complemented by developmentally appropriate activities that reinforce aspects of story comprehension skills such as joint Readers Theatre and puppet play, storyboards, pair-share activities, dictated stories, and storytelling.

Finally, parallel to assessment, narrative comprehension instruction can be targeted for specific groups of children (e.g., children with poor compre-

hension who need extra help or poor decoders who need instruction without a decoding barrier). Instruction in narrative comprehension of picture books might be especially helpful for bilingual or remedial programs because it emphasizes building comprehension skills when decoding skills are not very sophisticated. If research confirms the instructional validity of the NC task for young children, then intervention programs in preschools and primary grades will have the potential to remedy some of the differences in narrative skills that children often bring with them to formal education. We also think that the tenets of the NC task can be translated to other genres and hope that future research shows how to assess and instruct young children about the structural elements and relations in many genres.

Caveats, wishes, and limitations

This set of studies was ambitious and meticulous. They required almost four years to complete, and inevitably we discovered many things along the way that we wish we had done differently. On the one hand, we are eager to point out the limitations of our work so that others may improve and extend this line of inquiry. Yet, on the other hand, we believe that the reported studies are an important beginning for improved assessments of young children's comprehension which have many implications for reading and education. In the following section, we respond to limitations of these studies noted by reviewers and us.

Language and narrative comprehension

The NC task was designed with pictures to assess children's thinking about stories in order to minimize the effects of decoding, vocabulary, listening, and memory that usually influence assessments of children's early reading and listening comprehension. However, we recognize that children, as they examine picture books, may label pictures and make up sentences, whether vocalized or not. In this manner, prior knowledge, prior book experience, vocabulary, and language development may all affect how children comprehend wordless picture books. That is why we need to temper the claims about the NC task to say that linguistic factors have been minimized, but we acknowledge that they may still operate in the task in various ways for children. It would have been useful to include measures of children's language development so that the contributions of language to NC success might have been estimated, and perhaps future research will do that. Our indi-

rect evidence suggests that NC scores correlated better with GMRT comprehension than vocabulary scores, but there were strong correlations among phonological tasks and NC scores. The simultaneous development of narrative and linguistic skills among 3- to 8-year-olds, coupled with the multicollinearity of the data, make it difficult to separate the contributions of language and intelligence to narrative comprehension development.

Assessing validity

Validating a new task such as the NC is always difficult because the usual practices involve correlating performance on a new task with conventional benchmarks. We are uncertain about other assessments as appropriate benchmarks for NC skills. We used the GMRT and the ITBS because they are accepted standardized tests of reading suitable for young children, but they correlated differently with NC scores. NC scores were correlated with concurrent measures of the GMRT but were good predictors of ITBS scores given a year later. Both standardized tests are dissimilar to the NC task in many ways, so the actual cognitive processes that are implicated in the correlations are not clear. Likewise, the QRI-II is a reasonable benchmark task for reading, but it only revealed concurrent validity and not predictive validity with NC scores. After the studies had been completed, we realized that it would have helped to administer a listening comprehension task to children to test concurrent and predictive validity.

At least three problems confound assessments of validity in this research and other studies. First, the degree of similarity in content and procedures varies widely among tasks and influences the degree of correlation observed. Second, the choice of a benchmark is usually an established task, such as a standardized reading test, but there is no reason to assume that NC skills should be or need be correlated with the skills assessed on a standardized reading test, especially given the multiple skills involved in reading. Narrative comprehension has strong construct validity, but there are few tasks to measure it, and none have attained the status of a benchmark task for reading. Third, many cognitive assessments given to children display multicollinearity because they may overlap in processes and format. More important, they occur simultaneously with many rapid development changes in childhood so they may be epiphenomena of other developmental changes. In the case of NC skills, it is difficult to partial out, regress, or subtract the influences of cognitive and linguistic development as measured by many other

potential tasks such as vocabulary, sequencing, or inferential reasoning.

Subject variables

It has been suggested that the picture books that were used to assess NC may have cultural biases because children from different cultures may create and understand narratives in different ways, with different emphases on elements and relationships. It is also possible that the pictures may be interpreted differently by children from various cultures. We acknowledge that this is possible, and future research can vary cultural backgrounds of subjects with the structure of diffent narrative tasks. However, in our studies, most children were from European American and African American families, and there were too few children from distinctive cultures to examine the potential for bias in our data. We believe that the potential for bias with school children in the United States is relatively small in the picture books we used, which were certainly comparable to other classroom reading materials, so the NC assessment appears reasonable for use with children who are familiar with American books and stories.

We also did not have enough older prereaders and very young readers to conduct complete analyses by reading ability. Thus, the degree to which NC skills correlate with reading ability is unclear. Our studies suggest that there is a relation, but it could be mediated by other factors. Future research may redress our low cell sizes and test the relationship with regression or Hierarchical Linear Modeling.

Task variables

The books that we created for the NC task were modified trade books, and some have suggested this may have influenced children in our studies. We think that using four different books in the various studies was a strong test of the replicability and generalizability of our research. However, we acknowledge that the black-and-white pictures may have been less interesting than more colorful books and that the spiral bindings are unlike regular books. The few deleted pages seemed irrelevant to us and also to the story lines, but future researchers can use authentic trade books to test our findings. We think a more important issue is how teachers can use the books in their classrooms for NC assessments. We believe that teachers can modify picture books, create picture books, or use regular trade books to assess children's narrative comprehension on a daily basis. The NC task provides one model for doing that, but teachers should be able to adapt similar questions to a variety of books so that informal assessment of comprehension becomes interwoven in the fabric of daily instruction.

We believe that reading is fundamentally a meaning-making process—during any encounter with text and pictures. A renewed focus on comprehension in early assessment and instruction can provide activities that are motivating, authentic, and enriching for young children's development into independent readers. Even children who cannot decode words can look at books, talk about books, construct meaning from books, and yield insights to teachers about cognitive abilities. It is crucial that young readers, from the first days of formal schooling, learn that reading is a meaning-making endeavor and that it is rewarding and fascinating, albeit effortful and sometimes frustrating. We hope that tools such as the NC task can foster classroom practices that help children become excited about books, understand the complex relationships among people and events in stories, talk about narrative events, and think and read more strategically about stories.

REFERENCES

ALLINGTON, R.L., & WALMSLEY, S.A. (1995). *No quick fix: Rethinking literacy programs in America's elementary schools*. New York: Teachers College Press.

ANDERSON, R.C., & PEARSON, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). New York: Longman.

ASTINGTON, J.W. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.

BERMAN, R.A., & SLOBIN, D.I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

BORNENS, M. (1990). Problems brought about by "reading" a sequence of pictures. *Journal of Experimental Psychology, 49*, 189–226.

BOTVIN, G.J., & SUTTON-SMITH, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology, 13*, 377–388.

BRANSFORD, J.D., & MCCARRELL, N.S. (1974). A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W.B. Weimer & D.S. Palermo (Eds.), *Cognition and symbolic processes* (pp. 189–229). Hillsdale, NJ: Erlbaum.

BRUNER, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

BURNS, M.S., GRIFFIN, P., & SNOW, C.E. (1999). *Starting out right*. Washington, DC: National Academy Press.

CASE, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Erlbaum.

CASE, R., & OKAMOTO, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*, 2.

CHALL, J.S. (1996). *Stages of reading development*. Fort Worth, TX: Harcourt Brace.

DICKENSON, D.K., & SMITH, M.W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly, 29*, 105–121.

DICKENSON, D.K., & SNOW, C.E. (1987). Interrelationships among prereading and oral language skills in kindergartners from two social classes. *Early Childhood Research Quarterly, 2*, 1–26.

DICKENSON, D.K., & TABORS, P.O. (1991). Early literacy: Linkages between home, school, and literacy achievement at age five. *Journal of Research in Childhood Education, 6*, 30–46.

DUKE, N.K. (2000). 3.6 minutes per day: The scarcity of information texts in first grade. *Reading Research Quarterly, 35*, 202–224.

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

FIESE, B.H., SAMEROFF, A.J., GROTEVANT, H.D., WAMBOLDT, F.S., DICKSTEIN, S., & FRAVEL, D.L. (1999). The stories that families tell: Narrative coherence, narrative interaction, and relationship beliefs. *Monographs of the Society for Research in Child Development*, 64(2).

FIVUSH, R., & HAMMOND, N.R. (1990). Autobiographical memory across the preschool years: Toward reconceptualizing childhood amnesia. In R. Fivush & J.A. Hudson (Eds.), *Knowing and remembering in young children* (pp. 223–248). Cambridge, England: Cambridge University Press.

FLOOD, J.E. (1977). Parental styles in reading episodes with young children. *The Reading Teacher*, 30, 864–867.

FOUNTAS, I.C., & PINNELL, G.S. (1996). *Guided reading*. Portsmouth, NH: Heinemann.

FRYE, D., & MOORE, C. (1991). *Children's theories of mind: Mental states and social understanding*. Hillsdale, NJ: Erlbaum.

GRAHAM, J. (1990). *Pictures on the page*. Sheffield, England: National Association for the Teaching of English.

HEATH, S.B. (1986). Separating "things of the imagination" from life: Learning to read and write. In W.H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 156–171). Norwood, NJ: Ablex.

LABERGE, D., & SAMUELS, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.

LESLIE, L., & CALDWELL, J. (1995). *The Qualitative Reading Inventory II*. Glenview, IL: Scott Foresman.

MANDLER, J.M., & JOHNSON, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111–115.

MCCABE, A., & PETERSON, C. (1991). *Developing narrative structure*. Hillsdale, NJ: Erlbaum.

MILLER, P.J. (1994). Narrative practices: Their role in socialization and self-construction. In U. Neisser & R. Fivush (Eds.), *The remembering self: Construction and accuracy in the self-narrative* (pp. 158–179). New York: Cambridge University Press.

MILLER, P., WILEY, A., FUNG, H., & LIANG, C.H. (1997). Personal storytelling as a medium of socialization in Chinese and American families. *Child Development*, 68, 557–568.

MORROW, L.M. (1985). Reading and retelling stories: Strategies for emergent readers. *The Reading Teacher*, 38, 870–875.

MORROW, L.M. (1990). Assessing children's understanding of story through their construction and reconstruction of narrative. In L.M. Morrow & J.K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 110–134). Englewood Cliffs, NJ: Prentice Hall.

NELSON, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 38(149).

NELSON, K., & HUDSON, J. (1988). Scripts and memory: Functional relationship in development. In F.E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes and individual differences* (pp. 147–167). Hillsdale, NJ: Erlbaum.

PARIS, S.G., & MAHONEY, G.J. (1974). Cognitive integration in children's memory for sentences and pictures. *Child Development*, 45, 633–642.

PARIS, S.G., WASIK, B.A., & TURNER, J.C. (1991). The development of strategic readers. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 609–640). New York: Longman.

PELLIGRINI, A.D. (1985). The relations between symbolic play and literate behavior: A review and critique of the empirical literature. *Review of Educational Research*, 55, 107–121.

PRESSLEY, M. (1996). More about the development of narrative skills. *Issues in Education*, 2, 69–72.

PRESSLEY, M., ALMASI, J., SCHUDER, T., BERGMAN, J., HITE, S., EL-DINARY, P.B., & BROWN, R. (1994). Transactional instruction of comprehension strategies: The Montgomery County, Maryland, SAIL program. *Reading and Writing Quarterly*, 10, 5–19.

RAND READING STUDY GROUP. (2001). *Reading for understanding: Towards an R&D program in reading comprehension*. Unpublished manuscript, Santa Monica, CA.

ROSER, N., & MARTINEZ, M. (1985). Roles adults play in preschoolers' response to literature. *Language Arts*, 62, 485–490.

ROTH, F.P. (1986). Oral narrative abilities of learning-disabled students. *Topics in Language Disorders*, 7, 21–30.

SCHMIDT, C.R., & PARIS, S.G. (1978). Operativity and reversibility in children's understanding of pictorial sequences. *Child Development*, 49, 1219–1222.

SCHMIDT, C.R., PARIS, S.G., & STOBER, S. (1979). Inferential distance and children's memory for pictorial sequences. *Developmental Psychology*, 15, 395–405.

SNOW, C.E. (1983). Language and literacy: Relationships during the preschool years. *Harvard Educational Review*, 53, 165–189.

SNOW, C.E., BURNS, M.S., & GRIFFIN, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

SNOW, C.E., & NINIO, A. (1986). The contracts of literacy: What children learn from learning to read books. In W.H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 116–138). Norwood, NJ: Ablex.

STALLMAN, A.C., & PEARSON, P.D. (1990). Formal measures of early literacy. In L. Morrow & J. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 7–44). Englewood Cliffs, NJ: Prentice Hall.

STEIN, N.L., & ALBRO, E.R. (1996). The emergence of narrative understanding: Evidence for rapid learning in personally relevant contexts. *Issues in Education*, 2, 83–98.

STEIN, N.L., & GLENN, C.G. (1979). An analysis of story comprehension in elementary school children. In R. Friedle (Ed.), *New directions in discourse processing*. Norwood, NJ: Ablex.

STEIN, N.L., & TRABASSO, T. (1982). What's in a story. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 213–267). Hillsdale, NJ: Erlbaum.

SULZBY, E. (1985). Children's emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly*, 20, 458–481.

TEALE, W.H. (1986). Home background and young children's literacy development. In W.H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 173–205). Norwood, NJ: Ablex.

VAN DEN BROEK, P. (2001). The role of television viewing in the development of reading comprehension (Center for the Improvement of Early Reading Achievement Archive Report No. 01-02). Ann Arbor, MI: University of Michigan.

VAN KRAAYENOORD, C.E., & PARIS, S.G. (1996). Story construction from a picture book: An assessment activity for young learners. *Early Childhood Research Quarterly*, 11, 41–61.

WELLMAN, H.M. (1990). *The child's theory of mind*. Cambridge, MA: Bradford Books/MIT Press.

WHITEHURST, G.J., & LONIGAN, C.J. (1998). Child development and emergent literacy. *Child Development*, 69, 848–872.

WILEY, A.R., ROSE, A.J., BURGER, L.K., & MILLER, P.J. (1998). Constructing autonomous selves through narrative practices: A comparative study of working-class and middle-class families. *Child Development*, 69, 833–847.

YUSSEN, S., & OZCAN, N.M. (1996). The development of knowledge about narratives. *Issues in Education*, 2, 1–68.

NC TASK PICTURE WALK

"We're going to look through this book together, and as we go through it I want you to tell me whatever you are thinking about the pictures or the story."

Picture Walk element	Score description	Score
1. Book-handling skills Orients book correctly, has sense of appropriate viewing speed and order (viewing errors may include skipping pages or speeding through pages)	Incorrectly handles book and makes more than two viewing errors	0
	Makes one to two viewing errors (i.e., skips pages)	1
	Handles book appropriately and makes no viewing errors	2
2. Engagement Behavioral and emotional involvement during Picture Walk, as judged by attention, interest in book, affect, and effort	Displays off-task behavior or negative comments	0
	Displays quiet, sustained behavior	1
	Shows several examples of attention, affect, interest, or effort (e.g., spontaneous comments)	2
3. Picture comments Discrete comments about a picture, which can include describing objects, characters, emotions, actions, and opinions as well as character vocalizations	Makes no picture comments	0
	Makes one picture comment or verbalization	1
	Makes two or more comments or verbalizations about specific pictures	2
4. Storytelling comments Makes comments that go across pictures and demonstrate an understanding that the pictures tell a coherent story—can include narration, dialogue, using book language and storytelling voice	Makes no storytelling comments	0
	Provides storytelling elements, but not consistently	1
	Through narration or dialogue, connects story events and presents a coherent story line	2
5. Comprehension strategies Displays vocalizations or behaviors that show attempts at comprehension such as self-corrections, looking back and ahead in book, asking questions for understanding, making predictions about story.	Demonstrates no comprehension strategies	0
	Exhibits one instance of comprehension strategies	1
	Demonstrates comprehension strategies at least two or more times	2

NC TASK PROMPTED COMPREHENSION QUESTIONS

Explicit questions

1. [Book closed, characters]
   Who are the characters in this story? (replacement words: people, animals)

2. [Book closed, setting]
   Where does this story happen? (replacement words: setting, take place)

3. [p. 10, initiating event]
   Tell me what happens at this point in the story. Why is this an important part of the story?

4. [p. 12, problem]
   If you were telling someone this story, what would you say is going on now? Why did this happen?

5. [p. 18, outcome resolution]
   What happened here? Why does this happen?

Implicit questions

1. [p. 6, feelings]
   Tell me what the people are feeling in this picture. Why do you think so?

2. [p. 8, causal inference]
   Why did the family get the robot?

3. [p. 16, dialogue]
   What do you think the people would be saying here? Why would they be saying that?

4. [p. 18, prediction]
   This is the last picture in the story. What do you think happens next? Why do you think so?

5. [Book closed, theme]
   In thinking about everything that you learned after reading this book, if you knew that your friend's dad was bringing home a robot for his family, what would you tell the dad to help him so that the same thing that happened in this story doesn't happen to him? Why would you tell him that? (replacement words: advice, warn)

NC TASK PROMPTED COMPREHENSION SCORING

The purpose of this Appendix is to provide a clear set of guidelines that describes what we did and what others should do if they use this task.

Rubrics for scoring the prompted comprehension questions

Explicit information

Characters
- 2 points = response indicates that characters are a family and a robot
- 1 point = response contains at least two of the story's characters
- 0 points = response provides only one character, or answer is inappropriate

Setting
- 2 points = response indicates an understanding of multiple settings
- 1 point = response provides only one setting
- 0 points = response is not an appropriate setting

Initiating event
- 2 points = response identifies the initiating event and links it with other relevant story information (e.g., with the problem)
- 1 point = response identifies the story element (e.g., the initiating event)
- 0 points = response fails to identify the initiating event

Problem
- 2 points = response identifies the problem and links it with other relevant story information (e.g., with the initiating action)
- 1 point = response identifies the story element (e.g., the problem)
- 0 points = response fails to identify the problem

Outcome resolution
- 2 points = response identifies the outcome resolution and links it with other relevant story information (e.g., the problem or the initiating action)
- 1 point = response identifies the story element (e.g., the initiating action)
- 0 points = response fails to identify the outcome resolution

Implicit information

Feelings
- 2 points = response indicates the inference of appropriate character feelings and connects the feelings to other pages or events
- 1 point = response indicates the inference of appropriate character feelings
- 0 points = response is not an appropriate inference of character feelings

Causal inference
- 2 points = response is an appropriate inference that is explained by using events from multiple pages
- 1 point = response is an appropriate inference that is derived at the page level
- 0 points = response fails to include an appropriate causal inference

Dialogue
- 2 points = response indicates the inference of appropriate character dialogue and connects the dialogue to other pages or events

(continued)

NC TASK PROMPTED COMPREHENSION SCORING (continued)

- 1 point = response indicates the inference of appropriate character dialogue
- 0 points = response does not concern character dialogue or is not relevant

Prediction
- 2 points = response represents a prediction that used previous action or pages from the story
- 1 point = response indicates a prediction that could be made based only on the last picture of the story
- 0 points = response does not contain an appropriate prediction

Theme
- 2 points = response indicates the incorporation of multiple events in order to create a narrative-level theme
- 1 point = response is a simple theme that uses information from one aspect of the story
- 0 points = response does not indicate an understanding of any theme

Examples

The following are examples of 0-, 1-, and 2-point responses to the initiating event question (A) and the prediction question (B). On the page for which the child is asked to describe the initiating event there is a picture of a girl pulling out the wires of the robot, which leads to the robot's becoming wild and ruining the house—the problem of the story. On the final page, for which the child is asked to infer a prediction, the father is fixing the robot, which suggests that the robot will be able to clean the house as it did when it was new.

0 points: Fails to identify element or make inference
A. "She's cleaning the robot. [This is important] because it's always nice to get cleaned, isn't it?"
B. "It's just the end. [I know this] because I don't see any more pages below it."

1 point: Picture-level responses
A. "The little girl is undoing all the cords and she's going to tie them into a bow so it looks like a girl. [This is important] maybe because she wants him to look more like a girl."
B. "It works again. [I know this] because they're fixing it."

2 points: Narrative-level responses
A. "The girl pulls out all of the wires. [This is important] because if we didn't know this, we wouldn't know why it was acting up."
B. "Maybe the machine tries to go away, but it gets caught by them. [I know this] because he's getting tired of doing all the chores.

MICHIGAN LITERACY PROGRESS PROFILE (MLPP)

1. Six scores were obtained from the three MLPP activities: one for each of the three tasks and three additional subscores derived from the phonemic awareness task.

2. The book used for assessing concepts about print possessed the features designated by the MLPP guidelines, including at least one example of (a) print and illustration on a single page, (b) multiple lines of text on a single page, and (c) a variety of punctuation marks. The book that we used was *Wake Up, Sun*, written by David L. Harrison and illustrated by Hans Wilhelm (1986, Random House). This book was labeled a "Step 1" book in the series Step Into Reading. Locations were marked in the book where the specific concepts about print questions would be asked (including the front, back, and individual inside pages).

3. Examples of each of the concepts about print categories: book concepts (e.g., front cover, back cover, title); reading concepts (print carries message, one-to-one match); directionality (e.g., beginning of text, left to right and top to bottom, return sweep); concept of word (e.g., first word, last word, word); concept of letter (e.g., first letter in word, last letter in word, one letter/two letters, letter names, capital letter, small letter); and punctuation marks (e.g., period, question, exclamation, quotation, comma).

4. In the rhyming section of the phonemic awareness task, the child was provided with a definition of rhyming and a few practice words. He or she was then given a series of eight words and asked to provide a word that rhymed with each item on the list (e.g., "Tell me a word that rhymes with *bat*"). If the child provided a nonsense rhyming word, he or she was asked, "Can you tell me another word that is a real word?" The child received 1 point for each correct rhyme; total rhyme scores could range from 0 to 8 points.

    For the phoneme blending section, the child received a definition of what it means to "put sounds together" and was allowed to practice with a few examples. Then he or she was presented with a series of sounds and asked to blend them into words. For example, the experimenter would say, "What word would I have if I put together the sounds /t/ /a/ /p/?" The child was asked to blend together eight examples of three-letter words, and 1 point was earned for each correct word blend. Phoneme blending scores could range from 0 to 8 points.

    The segmentation section assessed whether the children could discern sounds in words clearly enough to reproduce them in print. The child first listened to an explanation of what it means to "stretch out a word by thinking about how many sounds you hear," and then received several practice items. He or she was asked, "What are the sounds?" for a list of eight words and received 1 point for each correct response. Again, scores for segmentation could range from 0 to 8 points.

    If the child completely missed three consecutive items on any of the three phonemic awareness tasks and appeared confused about what was being assessed, the task was discontinued and the child was advanced to the next task in order to ensure that his or her experience was a positive one.

5. The following is the two-sentence "story": "I see a white cat in the sun. It is looking for some big toys." This is the kindergarten-level story, which all children received because they had all been identified as nonreaders. These sentences were read to the child at a normal pace and then dictated again very slowly so that the child could write down the words as they were heard.