JSLHR

Research Article

Growth of Finiteness in the Third Year of Life: Replication and Predictive Validity

Pamela A. Hadley,[a] Matthew Rispoli,[a] Janet K. Holt,[b] Colleen Fitzgerald,[a] and Alison Bahnsen[a]

Purpose: The authors of this study investigated the validity of tense and agreement productivity (TAP) scoring in diverse sentence frames obtained during conversational language sampling as an alternative measure of finiteness for use with young children.
Method: Longitudinal language samples were used to model TAP growth from 21 to 30 months of age for 37 typically developing toddlers. Empirical Bayes (EB) linear and quadratic growth coefficients and child sex were then used to predict elicited grammar composite scores on the Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) at 36 months.
Results: A random-effects quadratic model with no intercept best characterized TAP growth, replicating the findings of Rispoli, Hadley, and Holt (2009). The combined regression model was significant, with the 3 variables accounting for 55.5% of the variance in the TEGI composite scores.
Conclusion: These findings establish TAP growth as a valid metric of finiteness in the 3rd year of life. Developmental and theoretical implications are discussed.

Key Words: grammar, finiteness, language development, growth modeling, assessment

The acquisition of tense and agreement during the third year of life has both theoretical and applied importance. Competing theoretical accounts differ as to the amount of abstract linguistic knowledge attributed to children and the extent to which individual morphemes are characterized as part of a grammatical system. Generative accounts attribute much grammatical competence to young children (Caprin & Guasti, 2009; Radford, 1990; Wexler, 1994, 1998; Yang, 2002, 2004), with emphasis placed on similarities between morphemes[1] in development. For example, Radford (1990) pointed out that a syntactically related variety of tense and agreement (T/A) morphemes emerged together shortly after the second birthday. Rice, Wexler, and Hershberger (1998) demonstrated similarity in growth trajectories for a subset of these morphemes for typically developing children from 3 to 6 years of age. However, there has been less systematic investigation of growth regarding morphemes between the ages of 2 and 3 years. Thus, linkages between the emergence of the finiteness system during the toddler years and its subsequent mastery in the preschool years have not been empirically established.

In contrast, researchers working within a usage-based perspective (Pine, Conti-Ramsden, Joseph, Lieven, & Serratrice, 2008; Theakston, Lieven, Pine, & Rowland, 2005; Wilson, 2003) have focused on the development of individual T/A morphemes in English-speaking children between the ages of 2 and 3 years. Usage-based accounts ascribe little abstract grammatical knowledge to young children, and therefore, systematicity in early development is not predicted. Usage-based accounts emphasize that children's first uses of T/A morphemes are often in high-frequency, lexically based constructions. These initial uses are rotes, which are not structurally represented as individual morphemes (MacWhinney, 1978). With development, children's constructions are said to "grow gradually in abstractness" (Tomasello, 2005, p. 316). Yet, we are not aware of empirical studies demonstrating a direct facilitative effect for rotes and limited scope formulae on the acquisition of a productive grammar. On the other hand, generative accounts have not dealt effectively with the phenomena of rote and limited scope formulae, and this too must be systematically explored.

[a]University of Illinois at Urbana–Champaign
[b]Southern Illinois University, Edwardsville

Correspondence to Pamela A. Hadley: phadley@illinois.edu

Alison Bahnsen is now at Des Moines Public Schools, IA.

Editor: Rhea Paul
Associate Editor: Stephanie Stokes

Received January 8, 2013
Revision received June 4, 2013
Accepted August 19, 2013
DOI: 10.1044/2013_JSLHR-L-13-0008

[1]We use the term *morpheme* to refer to the affixes third-person singular present –3s and past *–ed*. We use the term *morpheme categories* for the? word-specific paradigms of copula BE, auxiliary BE, and auxiliary DO. The copula BE (as an example) is not a single morpheme but a paradigm of six words: *am, is, are, was, were, be*. Technically, each of these words is a morpheme in and of itself.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

During the earliest period of grammatical development, it is possible that grammatical knowledge may be overestimated by traditional measures of accuracy for children whose uses of T/A morphemes are contained primarily in high-frequency constructions.

Gradual morphosyntactic learning (GML) is an alternative account of the development of tense and agreement (Rispoli & Hadley, 2011; Rispoli, Hadley, & Holt, 2009, 2012). GML combines insights from maturational accounts (Radford, 1990; Rice et al., 1998; Wexler, 2003), variational learning (Yang, 2002) and other generativist accounts of grammatical development (Pinker, 1984) with models of adult sentence production (Bock 1982; Bock & Levelt, 1994; Lapointe & Dell, 1989). GML assumes that children have knowledge at their disposal to organize language input; that is, the principles of universal grammar (UG) give children the tools to build grammar (Pinker & Jackendoff, 2009). GML assumes that children's knowledge of the distinction between predicate versus argument versus adjunct, as well as the principle of structure dependence, guides their learning of phrasal and clausal relationships (Crain & Nakayama, 1987; Rispoli, 1994; Van Valin, 2009). GML also assumes that UG endows children with knowledge of the grammatical features that are relevant to building a grammar (Pinker, 1984; Slobin, 1985). Children learn which features are active in the language from the input. By using their knowledge of clausal structure and structurally dependent representations, children can then relate morphemes that appear in diverse syntactic contexts in the input to one another because they are in complementary distribution. Sensitivity to this distributed evidence forms the basis of morphosyntactic learning and unifies morphemes with distinct surface forms through shared grammatical features (e.g., copula *is,* verb affix –3s; auxiliary *did,* verb affix *–ed*; cf. Hadley, Rispoli, Fitzgerald, & Bahnsen, 2011; Rispoli et al., 2012).

GML also recognizes two production modes: direct activation and grammatical encoding (Rispoli & Hadley, 2011). The difference between the modes has implications for the measurement of finiteness growth. *Direct activation* is a production mode that results from "direct associative connections between referential content and phonetic content" (Bock, 1982, p. 24). This route does not employ true morphosyntactic representations. Direct activation is thought to arise from high-frequency word and morpheme combinations in input and use (Bybee, 2006). Stereotypic multimorphemic expressions (e.g., *What's that? I don't know.*) and limited scope formulae with invariant frames (e.g., *it's a X*) are likely to be produced by direct activation, bypassing grammatical representations. GML recognizes the influence of input frequency in shaping children's early rote and formulaic productions. However, GML proposes that measurement of young children's knowledge of T/A morphemes should minimize the influence of these productions to best reflect grammatical development (Rispoli et al., 2009).

The second production mode is grammatical encoding. *Grammatical encoding* refers to the real-time process of assembling sentences from grammatical representations (Bock & Levelt, 1994). The clausal representations and grammatical features associated with these representations have been formed gradually through the process of UG-guided implicit learning. As children identify the grammatical features active in the language, the strength of those features increases in sentence production, leading to the production of more diverse sentences with surface forms encoding those features. The representations are not tied to individual lexical items, but in order to be inserted in sentences, lexical items must be compatible with the grammatical features generated from these representations. Thus, children's production of T/A morphemes in diverse, low-frequency sentence frames (e.g., *the kitty likes milk*) are viewed as better indicators of young children's grammatical knowledge because these combinations are more likely to be produced via grammatical encoding than by direct activation.

We developed and evaluated a tense agreement productivity (TAP) score (Hadley & Short, 2005; Rispoli et al., 2009) grounded in GML's psycholinguistic framework to measure the onset of finiteness in young children. The TAP score combines the uses of five T/A morpheme classes into a single score, consistent with our theoretical assumption that these morphemes are unified by clause representation and shared grammatical features. The score is type based, requiring that these uses be produced across sufficiently different contexts. The scoring criteria also act as a high-frequency filter. Together, these scoring decisions are intended to reduce the influence of forms that are more likely to be produced via direct activation and to emphasize forms that are more likely to be produced by grammatical encoding.

Increases in the TAP score are interpreted as reflecting increases in the strength of grammatical features in sentence production. Using this measurement approach, Rispoli et al. (2009) revealed developmental growth in finiteness from 21 to 30 months of age. Growth modeling indicated that a no-intercept quadratic model was the best fit to the data. Finiteness marking was essentially absent at 21 months, with statistically significant linear growth occurring at 21 months and accelerating between 21 and 30 months for the group as a whole. The average trajectory was characterized by T/A morphemes growing at a rate of 0.58 morphemes per month, with additional acceleration of 0.07 morphemes per month between 21 and 30 months. The average growth trajectory corresponded to approximately two sufficiently different uses at 24 months of age and 11 uses at 30 months of age.

Significant variation between children was also observed in the linear and quadratic growth rates in the Rispoli et al. (2009) study. Children's individual linear and quadratic growth coefficients accounted for 45% of the variance in accuracy of finiteness marking (i.e., percentage correct in obligatory contexts) 3 months later. Although Rispoli et al. provided preliminary evidence for the validity of TAP growth by linking children's rate of growth with subsequent accuracy, the measures of TAP growth and accuracy were both obtained from conversational language sampling. A stronger test would involve the use of standard probes that control the obligatory contexts for the finiteness morphemes when computing accuracy outcomes.

In this study, we explored the influence of child sex on the early growth of finiteness. Although a seminal meta-analysis by Hyde and Lynn (1988) did not reveal evidence for sex differences in the verbal abilities of children and adults, this meta-analysis excluded studies of toddlers between the ages of 18 months and 3 years (see Hyde & Linn, 1988, p. 55). Studies conducted in early childhood may show a female advantage (Bornstein, Hahn, & Haynes, 2004). For example, sex differences have been documented for vocabulary growth during the first and second years of life, with girls showing earlier onset and more rapid vocabulary growth than boys (Bauer, Goldfield, & Reznick, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). Girls' vocabulary size advantage is also evident in the cross-sectional normative data for the MacArthur-Bates Communicative Development Inventories (MCDI; Fenson et al., 2007). Sex differences have also been documented for the emergence of word combinations at age 2 years, with boys being more likely than girls to experience late language emergence (Zubrick, Taylor, Rice, & Slegers, 2007). Given that the current study covered the period from 21 to 36 months, the exploration of child sex seemed appropriate.

The purpose of this study was to evaluate the reliability and validity of our type-based approach for assessing early growth in finiteness and to establish developmental expectations for the growth of this system. Our first objective was to determine whether the TAP growth model from 21 to 30 months of age characterized by Rispoli et al. (2009) would replicate in a new and larger sample, thereby demonstrating the reliability of the growth expectations between two different samples. Our second objective was to determine whether children's TAP growth coefficients would predict accuracy of finiteness marking on a standardized set of probes at 36 months of age. Together, converging evidence would indicate that it is possible to characterize early growth trajectories for the acquisition of finiteness by applying TAP scoring criteria to conversational language samples with young children, who are often challenging to assess with formal measures.

Method

Participants

A total of 37 typically developing toddlers (18 male, 19 female) contributed data for modeling longitudinal growth in finiteness. All of the toddlers were required to demonstrate (a) a passing score on the communication section of the Ages and Stages Questionnaires (ASQ; Bricker & Squires, 1999) at 21 and 24 months of age, (b) expressive vocabulary abilities at or above the 10th percentile at 24 months[2] (Fenson et al., 2007), and (c) production of at least two different words with each of the final consonants /t, d, s, z/ in their spontaneous language samples by 30 months (Bleile, 1995). A total of 32 toddlers (17 male, 15 female) were included in the outcome analysis. To be included in the outcome analysis, children were also required to pass a phonological screening for /t, d, s, z/ in nonmorphophonemic contexts at 36 months (Rice & Wexler, 2001). From the original 37 children, four were lost to attrition after the 30-month visit, and one was not compliant with formal testing at 36 months. Informed consent was obtained from the children's parents, and the children assented to participation in play sessions at each measurement point and to formal testing at 36 months.

The children were reported by their parents to be White (n = 30), Black (n = 4), or biracial (i.e., White/Black [n = 3]), and all were reported to be non-Hispanic. All of the families spoke Standard American English. Maternal educational levels included completion of high school (n = 2), associate's degree or some college (n = 5), bachelor's degree (n = 17), and advanced degree (n = 13).

Procedure

Longitudinal data were obtained at six measurement points, when children were approximately 21, 24, 27, 30, 33, and 36 months of age. Parent report checklists were mailed to families before the first four visits. The ASQ was used at 21 and 24 months of age to ensure that the children were developing typically in the area of communication and to screen for difficulty in fine motor, gross motor, social, and cognitive development. Parents also completed the Words and Sentences subtest of the MCDI at 21, 24, 27, and 30 months in order to provide an indication of developmental changes in their children's expressive vocabulary abilities.

Spontaneous language samples were gathered in a lab playroom. Three sound-field microphones were used to produce high-quality digital recordings. In addition, the children were encouraged to wear a remote (radio transmitter) lapel microphone in a vest. All inputs were mixed through a mix-pad and were burned to CDs and DVDs.

Each 1-hr session was divided into two sampling contexts. The first sampling context included 30 min of parent–toddler free-play with age-appropriate toys. The parents were encouraged to play with their children "as they would at home." For the second sampling context, an examiner joined the parent–toddler dyad to introduce semistructured play scenarios and to shape the conversational discourse in a way that created more opportunities for the children to use T/A morphemes. The goal of the second sampling context was to ensure that the absence of T/A morphemes was not due to a lack of conversational opportunities. Examiners structured the discourse with the goal of creating at least three opportunities for each of the target T/A morphemes. If a child produced sentences with several obligatory contexts for one or two morphemes during the first 30 min of parent–toddler play, the examiner focused on creating opportunities for the remaining morphemes in the second 30 min. By 30 months of age, the average number of obligatory contexts for the T/A morphemes in child

[2]There were two girls whose MCDI scores fell below the 10th percentile at one measurement point between 21 and 30 months: one girl at 21 months and the other at 30 months.

sentences was third-person singular (–3s), 11.59; past tense *–ed,* 2.81; copula BE 53.24; auxiliary BE 17.43; and auxiliary DO 11.46.

The examiners were trained to talk about the states, actions, and properties of the toys in the play environment and to introduce lexical nouns as grammatical subjects in order to create more opportunities for scorable uses of copula and auxiliary BE (cf. Oetting & Hadley, 2009). For example, the baby doll caregiving routine was expected to create opportunities for copula BE (e.g., *Nina is/The babies are hungry/cold/tired*) and for –3s (e.g., *The baby needs a bath*). Windup toys were used to create opportunities for *3s* past tense *–ed,* and auxiliary BE (e.g., *The bunny hops/hopped/is hopping*). Blowing bubbles was used to create opportunities for past tense *–ed* (e.g., *The bubble popped.*) Other verbs such as *float, land, stomp,* and *step* were also modeled to encourage opportunities for regular past tense. Puzzles and Mr. Potato Head toys created opportunities for children to use –3s on state verbs (e.g., *The lion goes here. He needs a hat.*). Discourse strategies were also used to create contexts for DO support. Examiners might falsely assert, *Mr. Potato Head needs another arm* to create an opportunity for *No, he doesn't* (cf. Cleave & Fey, 1997). Strategic questions such as *Who wants pizza?* were also posed to create opportunities for elliptical responses such as *I do!* or *The bear does.*

The examiners adjusted their discourse to the children's developmental levels. Although some children produced numerous sentences with obligatory contexts for T/A morphemes, other children needed more discourse support. For example, to create opportunities to use –3s while playing with a windup toy, the examiner might say *The seal spins* followed by a question (e.g., *What else does he do?).* The examiners could also use production prompts and intonation cues (e.g., *Tell me what the elephant does. He~*) to support independent production of sentences as the conversation continued (e.g., *He spin* or *He spins*). However, some children's grammatical development was not sufficiently advanced for T/A morpheme production until 27 to 30 months of age.

Transcription was completed in a series of passes by trained transcribers. All of the sessions were transcribed following the conventions of the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2010) software program and the general procedures described in the authors' previous publications (Hadley & Holt, 2006; Hadley & Short, 2005; Rispoli, 2005; Rispoli et al., 2009). The first transcriber completed an initial pass, listening to child utterances no more than three times, so that differences in the children's general intelligibility were realistically reflected in the transcripts. A second transcriber completed a consensus pass (cf. Shriberg, Kwiatkowski, & Hoffmann, 1984), relistening to the recording and marking as unintelligible any part of the transcript that could not be confirmed. Consensus transcribers also removed T/A morphemes that could not be confirmed. If the consensus transcriber noted a T/A morpheme that was not originally transcribed, a third transcriber listened to the specific utterance to confirm the presence of the morpheme before it was inserted into the transcript.

Once the consensus transcript was completed, we added utterance codes to exclude imitations, exact immediate self-repetitions, and routine expressions (e.g., counting, lyrics) from further analyses. Spellings of words were standardized to ensure that general measures of lexical diversity were not inflated (e.g., *ducky* and *duck* were counted as a single word; Hadley et al., 2011). All uses of bound morphemes (e.g., *jump/ed*) and contractions (e.g., *He/'s*) as well as omissions of bound morphemes (e.g., *It go/*3s there*) and closed class words (e.g., *Horse *is run/ing*), were coded using standard SALT conventions. Additional codes were inserted to identify all obligatory contexts for copula BE (e.g., *that guy/'s*[c] *mine.*), auxiliary BE (e.g., *are*[a] *you come/ing*?), and auxiliary DO (e.g., *it does*[d] *not fit*.).

At 36 months, the Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) was administered to the children with their parents present. The parents were informed that the test was designed for children 3 to 8 years of age and that some children may have trouble producing responses. If parents prompted responses or rephrased a prompt, they were reminded of the examiner's need to follow the test script.

The TEGI probes created 54 obligatory contexts for T/A morphemes, resulting in summary scores for Third Person Singular, Past Tense, Copula/Auxiliary BE, Auxiliary DO, and an Elicited Grammar Composite. The TEGI probes were administered on two different days within 1 to 2 weeks of each other.[3] On the first day, the children completed the phonological screener, the third person singular probe, and the past tense probe. On the second day, the BE/DO probe was administered. Given the children's young age and the potential for young children to experience fatigue during formal testing, we made some minor modifications to the TEGI manual's reprompting instructions and trial items. For example, the manual states that if the children produce an irregular for a regular verb (i.e., *went* for *climbed*) or vice versa (*hammered* for *made*), the children should be reprompted for the correct type of verb. However, either type is scorable. Thus, if the children were moving slowly or with difficulty through the probes, we accepted the first scorable response. Because children had more difficulty "asking the puppet" questions on the BE/DO probe, we provided three additional trial items to model questions. When the children were reluctant to produce questions, the parents participated in the trial items to encourage child participation. Several children (n = 13) remained reluctant to produce DO questions on the BE/DO probe. For these children, the examiner provided the child's turn in the script, and only responses to BE attempts were scored. All of the child responses were transcribed for subsequent scoring and analysis.

[3]A third session was scheduled for one child who experienced fatigue and subsequent noncompliance during the TEGI administration.

Measures

TAP scores (Hadley & Short, 2005; Rispoli et al., 2009) were obtained from the 1-hr spontaneous language samples. The TAP score is designed to protect against the undue influence of rotes and limited scope formulae that were produced via direct activation (Hadley & Short, 2005; Rispoli et al., 2009). The TAP score has been shown to be a good indicator of a child's developmental level (Guo, Van Horne, & Tomblin, 2011) and to differentiate typically developing preschool children from preschoolers with specific language impairment (SLI; Gladfelter & Leonard, 2013). The TAP score filters out uses of T/A morphemes that are potentially produced by rote early in development. Irregular forms (e.g., *went, came, has*) are not counted. Similarly, uses of *don't* and *ain't* are excluded because these negative forms involve irregular changes from the affirmative forms. Uses of copula and auxiliary forms contracted to pronouns are also excluded because combinations of pronouns with contracted copula, auxiliary BE and DO (e.g., *he's, it's, what's, where's, here's*) are high frequency in the input and are packaged together as a single syllable (Li & Shirai, 2000; MacWhinney, 2000). This also excludes uses of T/A morphemes in common limited scope formulae such as *it's a X, that's a X,* and *where's X.* Finally, the type-based nature of the TAP score limits the influence of repetition. Repetitions of verb inflections with the same lexical verbs and repetitions of the same grammatical subject–copula/auxiliary combinations are not counted.

To compute a noncumulative TAP score, one point is earned for each sufficiently different use of five T/A morphemes up to a maximum score of 5 for each class (i.e., –3s, past tense *–ed,* copula BE, auxiliary BE, and auxiliary DO) at each measurement point. TAP scores can range from 0 (i.e., nonemergence of T/A morphemes) to 25 (i.e., diverse uses across five morpheme classes). Uses of –3s and *–ed* are sufficiently different if they appear on different verb stems. Overregularizations of –3s and *–ed* (e.g., *have-3s*, *go ed, come ed*) on irregular verb stems are also counted because overregularization provides strong evidence of an affix's productivity. Uses of copula/auxiliary forms are sufficiently different if they are distinct forms (e.g., *do* vs. *does; is* vs. *are*) or if the same form appears with different subjects (e.g., *Baby is hungry* vs. *She is hungry*). For copula/auxiliary forms, a different type is identified by the combination of grammatical subject and finiteness morpheme. Thus, *What is this?* and *This is a dog* are the same type, but *Does it go here?* and *Did it fit?* are different types. To be scored, finiteness morphemes were required to be used in a sentence with an expressed grammatical subject and marked correctly for tense and agreement. In other words, markers used without a subject (e.g., *goes here*) or in error (e.g., *they needs help*) did not contribute to the TAP score. Multiple forms within a single sentence were counted if the subject had scope over conjoined verb phrases (e.g., *It goes in the pan and cooks*). In sum, the combination of type-based criteria for diverse contexts and a high-frequency filter increases the likelihood that developmental change in TAP scores reflects the strengthening of grammatical features in sentence production, which is interpreted as growth in children's emerging knowledge of finiteness.

The TEGI Elicited Grammar Composite was the outcome measure. Computation of the composite score required at least three of the four subscores (i.e., –3s, past tense *–ed,* BE and DO). Accuracy scores for each subscale were then summed and divided by the number of subscores. For this study, a minimum of three scorable responses was required to compute a subscore. Scoring of individual items followed the instructions in the manual, supplemented by the following decisions: On the third-person singular probe, *hit* and *put* were accepted as incorrect responses if there was no reason to think that the child was attempting a past form (i.e., if the child had not used another past form during the probe or a past adverbial such as *yesterday*), and on the BE/DO probe, pronoun reference errors were considered unscorable (e.g., *They are* for a *She is* target). TEGI composites could range from 0% to 100%.

Reliability

Transcribers completed approximately 20 hr of training. Before transcribing the data used in the analyses, the transcribers were required to complete three consecutive training transcripts at 80% reliability overall and 90% reliability for T/A morphemes. Transcription reliability was obtained through consensus procedures designed to reduce measurement error and to err on the side of conservative transcription whenever possible (Shriberg et al., 1984). Consensus procedures were adopted given the difficulty inherent in transcribing very young children's speech, consistent with the procedures used by other investigators to transcribe toddlers' speech (e.g., Eisenberg, Guo, & Germezia, 2012; Rescorla, Roberts, & Dahlsgaard, 1997; Theakston & Rowland, 2009). These procedures ensured that two individuals agreed on the transcription of all T/A morphemes produced in more than 42,000 complete and intelligible child utterances.

To determine the interrater reliability of the scoring procedures, we randomly selected four children's TAP scores at each measurement point and scored them independently. Scoring agreements and disagreements were identified for every scoring decision in the 16 samples. For example, scorers should agree that *It is hot* should be scored and that *It's hot* should not be scored. Agreement occurred when both scorers identified the same number of instances to score. Out of 456 uses of T/A morphemes, there were only three disagreements between the original decisions and the independent scorer's decisions, with agreement for individual samples ranging from 96% to 100% (M = 99.5, SD = .01).

To determine the interrater reliability of TEGI scoring, we randomly selected and retranscribed four children's TEGI recordings. Reliability was computed for each probe for each child. For the third person singular probe and the past tense probe, transcription accuracy and scoring decisions on verb forms were examined. For the BE/DO probe, transcription accuracy scoring decisions on all subject and verb forms were considered. Agreements were divided by total decisions and multiplied by 100. Agreement for the individual probes ranged from 81% to 93% and from 83% to 94% across participants.

Growth Models

Hierarchical linear modeling (HLM) was used to analyze developmental change in the TAP scores (Holt, 2008: Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011b). HLM allows us to investigate what type of linear or curvilinear model best describes individual children's change over time in their TAP score and how these growth trajectories vary across children (i.e., Does one TAP growth pattern fit all children equally well, or is there substantial variability in the growth patterns among the children?). When applied to the study of longitudinal change, HLM involves two levels of analysis: (a) an individual growth model that represents changes in each child's score over time and (b) a between-child model that represents differences among children in their growth trajectories.

A series of growth models was used to explore within-individual change over time in the TAP scores. Full information maximum likelihood (FML) estimation was used to compute a deviance statistic to determine the most appropriate growth model. Although FML can result in biased estimates, when the number of regression parameters is small, this effect is minimal (Raudenbush & Bryk, 2002, p. 53).

Model 1 explored TAP growth as a linear function of age centered at 21 months (see Equation 1):

$$\begin{aligned} \mathrm{TAP}_{ti} &= \pi_{0i} + \pi_{1i}(\mathrm{age}_{ti} - 21) + e_{ti} \\ \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} + r_{1i} \end{aligned} \quad , \tag{1}$$

where TAP_{ti} is the observed TAP score for child i at t months, and e_{ti} is the deviation of child i from his or her growth trajectory at time t. The e_{ti} are assumed to be normally distributed, with a mean of 0 and variance σ^2. The parameter π_{0i} represents the status of child i at the centering point of 21 months, and π_{1i} represents the linear growth rate for child i. These two parameters reflect characteristics of growth in each child's TAP score. The individual growth coefficients, πs, are comprised of both fixed and random components, represented in the level 2 model as β and r, respectively. The fixed component β reflects the average growth parameter for the group. The random component r is the residual, or the individual's deviation from the fixed component. For instance, a fixed-effects linear coefficient is the estimated linear growth for all children in the population, whereas the random effect of the linear growth equation indicates how an individual child's linear growth differs from the population as a whole. The equation was age-centered at 21 months because in a previous study (Rispoli et al., 2009), we found that few children used T/A morphemes at this age, and therefore we expected to capture a true 0 at 21 months. If this finding was replicated, centering at 21 months would allow us to simplify the model (Model 1a) by eliminating the intercept (see Equation 2). A likelihood ratio test was conducted to determine if the fit was significantly worse after eliminating the intercept (i.e., comparing the fit of Model 1 to Model 1a; Holt, 2008).

$$\begin{aligned} \mathrm{TAP}_{ti} &= \pi_{1i}(\mathrm{age}_{ti} - 21) + e_{ti} \\ \pi_{1i} &= \beta_{10} + r_{1i} \end{aligned} \tag{2}$$

Model 2 was expanded to include quadratic growth across time (see Equation 3) by adding time squared as a second predictor in the model. The intercept was dropped, pending the outcome of the deviance test. In this model, the linear growth parameter π_{1i} is the linear rate of change for individual i at the centering point of 21 months. The quadratic growth parameter π_{2i} reflects the curvature or acceleration/deceleration in each child's overall growth across time. Larger positive values of π_{2i} imply more rapid growth in TAP score. Likewise, the fixed effects β_{10} and β_{20} denote the overall linear growth at 21 months, and overall quadratic change from 21 to 30 months, respectively. To determine if a quadratic model was a better fit to the data than a linear model, we conducted a likelihood ratio test comparing the deviance between Model 1a and Model 2.

$$\begin{aligned} \mathrm{TAP}_{ti} &= \pi_{1i}(\mathrm{age}_{ti} - 21) + \pi_{2i}(\mathrm{age}_{ti} - 21)^2 + e_{ti} \\ \pi_{1i} &= \beta_{10} + r_{1i} \\ \pi_{2i} &= \beta_{20} + r_{2i} \end{aligned} \tag{3}$$

Sex differences in growth parameters were tested in Model 3 by adding sex as a time-invariant predictor at level 2 (Equation 4). The parameters β_{11} and β_{21} capture the sex differences in linear growth at 21 months and in the acceleration/deceleration over time, respectively.

$$\begin{aligned} \mathrm{TAP}_{ti} &= \pi_{1i}(\mathrm{age}_{ti} - 21) + \pi_{2i}(\mathrm{age}_{ti} - 21)^2 + e_{ti} \\ \pi_{1i} &= \beta_{10} + \beta_{11}(\mathrm{sex})_i + r_{1i} \\ \pi_{2i} &= \beta_{20} + \beta_{2i}(\mathrm{sex})_i + r_{2i} \end{aligned} \tag{4}$$

In a second phase of analysis, the empirical Bayes (EB) growth coefficients, πs, were generated for the TEGI analyses. The EB coefficients are estimates of the child-level growth parameters that use information from the individual child and from similar children in producing optimal growth parameter estimates (Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011a). The TEGI composite scores at 36 months were regressed on the growth coefficients, controlling for child sex, to examine the predictive validity of the growth coefficients on the accuracy of T/A marking at 36 months.

Results

Descriptive statistics by measurement point are reported in Table 1. The average number of child utterances increased from 161 at 21 months to 387 at 30 months in 1 hr

Table 1. Descriptive statistics for the developmental measures by measurement point.

Measure	Minimum	Maximum	Mean	*SD*
MCDI Total Words score				
21 months	37	586	198.54	134.26
24 months	102	640	352.95	157.30
27 months	140	668	483.95	140.01
30 months	275	675	561.84	107.05
C&I utterance				
21 months	12	516	161.89	105.87
24 months	75	540	256.46	102.38
27 months	174	601	337.70	114.83
30 months	182	663	387.19	126.96
MLU				
21 months	1	2.46	1.34	0.36
24 months	1.12	3.94	1.86	0.59
27 months	1.33	4.09	2.46	0.65
30 months	1.58	4.62	2.93	0.68
TAP score				
21 months	0	4	0.46	1.07
24 months	0	10	2.11	2.90
27 months	0	19	5.92	5.12
30 months	1	24	9.81	6.39
TEGI score				
36 months	6.67	97.44	54.37	27.37

Note. MCDI = MacArthur–Bates Communicative Development Inventories; C&I utterance = total number of spontaneous, complete, and intelligible utterances; MLU = mean length of utterance in morphemes from the first 30 min of parent–child interaction; TAP score = tense agreement productivity score; TEGI = Test of Early Grammatical Impairment.

of conversational language sampling. The number of utterances increased steadily between 21 and 27 months and then leveled off. The average mean length of utterance (MLU) in morphemes from the first 30 min of parent–child interaction increased from 1.34 to 2.93 from 21 to 30 months. Similar developmental changes were observed during the second 30 min of semistructured play with the examiner, with average MLUs increasing from 1.38 to 2.80 between 21 to 30 months.

The TAP scores showed a general developmental pattern of increasing mean scores and increasing variation over time. Mean scores increased steadily with age, from a mean of 0.46 (SD = 1.07) at 21 months to 9.81 (SD = 6.39) at 30 months. At 21 months, 27 children (73%) had a score of 0, and eight (22%) had very low scores (i.e., 1 to 3). These scores indicate very minor variability in TAP scores at the initial measurement point. Variation was also evident in the 36-month TEGI composite scores (M = 54.37, SD = 27.37).

An unconditional means model was generated first to examine the variance in TAP scores both between and within children. The unconditional means model (Model 0) resulted in a significant intercept, $b_{00} = 4.62$, $p < .001$, 95% confidence interval (CI) [3.52, 5.72], and substantial intraclass correlation (ICC), $\rho_{ICC} = .157$, indicating that 15.7% of the variance in TAP scores was among children and 84.3% of the variation in TAP scores was within children. Growth parameters were entered into the model to determine if the within-child variability was random in nature or consistent with a development growth pattern.

A linear growth trajectory was examined in Model 1 (see Equation 1). The estimates indicated that the fixed linear coefficient and the between-child variability in linear growth were both statistically significant (see Table 2). That is, the average linear increase in TAP score was significantly different from 0, $b_{10} = 1.04$, $p < .001$, 95% CI [0.82, 1.26], with a growth of 1.04 morphemes per month, with significant variability in linear growth among children (i.e., $r_1 = .33$). However, the intercept was no longer significantly different from 0, $b_{00} = -0.27$, $p = .41$, 95% CI [–0.90, 0.36], indicating nonemergence of T/A morphemes at 21 months. A likelihood ratio test indicated that there was no improvement in fit between Model 1 and a more parsimonious model without the intercept, Model 1a, $\chi^2(3) = 3.81$, $p = .28$; therefore, the intercept was removed from the model (see Equation 2), as in Rispoli et al. (2009). When the linear model with no intercept was fit to the data, significant heterogeneity of variance in level 1 errors across children remained, possibly indicating that another growth parameter(s) was needed to account for this unmodeled variance.

Next, a quadratic model, Model 2 (see Equation 3), was fit to the data to model any change in linear growth that might be occurring over time. A likelihood ratio test indicated that the quadratic model was a significantly better fit to the data than the linear model, $\chi^2(3) = 30.22$, $p < .001$. Also noteworthy, the heterogeneity of variance in the residuals was no longer significant in the quadratic model. Through this model-building approach, a quadratic model with no intercept was deemed the optimal model to characterize TAP growth over time.

The quadratic model had a statistically significant linear term, indicating that the linear growth rate at 21 months was significantly different from 0, $b_{10} = 0.64$, $p = .006$, 95% CI [0.21, 1.07], and the quadratic trajectory from 21 to 30 months was also statistically significant, $b_{20} = 0.05$, $p = .043$, 95% CI [0.003, 0.09] (see Table 2). Together, these coefficients describe a growth pattern in which the rate of growth in TAP scores at 21 months was .64 morphemes per month, and this rate accelerated at a rate of .05 morphemes per month, resulting in an overall trend that was accelerating over time. However, there was also significant variability in this trend among children. Both the linear rate of growth in TAP scores at 21 months as well as the quadratic growth trend over time displayed significant between-child variation, $\tau_{11} = 1.46$, $p < .001$, and, $\tau_{22} = 0.01$, $p < .001$, respectively. Therefore, children had varying initial rates of TAP growth and varying TAP trajectories over time.

The significant random variability in linear and quadratic growth is indicated by the variation in the growth patterns (see Figure 1). Interestingly, there appear to be different growth patterns that lead to higher scores at 30 months. Three children with high 30-month predicted TAP scores and different growth patterns are highlighted with bold lines in Figure 1. In the figure, Line A represents the TAP growth of a child who has a strong linear growth rate at 21 months and then continues to increase in a linear fashion, Line B indicates a child with a strong initial linear rate of growth that decelerates somewhat over time, and Line C

Table 2. Fixed and random effects (and standard errors) for the growth models.

Variable	Model 0	Model 1	Model 1a[a]	Model 2[b]	Model 3
Fixed effect					
Intercept, β_{00}	4.62*** (0.56)	–0.27 (0.32)			
Linear growth, β_{10}		1.04*** (0.11)	1.00*** (0.11)	0.64** (0.22)	0.66* (0.31)
Sex					–0.04 (0.44)
Quadratic growth, β_{20}				0.05* (0.02)	0.06† (0.03)
Sex					–0.02 (0.04)
Random effect					
Intercept, r_0	4.98**	0.33			
Linear growth, r_1		0.33***	0.43***	1.46***	1.46***
Quadratic growth, r_2				0.01***	0.01***
Level 1 error, σ^2	26.70	4.79	4.97	2.53	2.52

Note. All of the models were age-centered at 21 months (n = 37).

[a]There was no significant difference in fit after removing the intercept from the linear model, $\chi^2(3) = 3.81$, $p = .28$. [b]There was a significant improvement in fit after adding the quadratic term, $\chi^2(3) = 30.22$, $p < .001$.

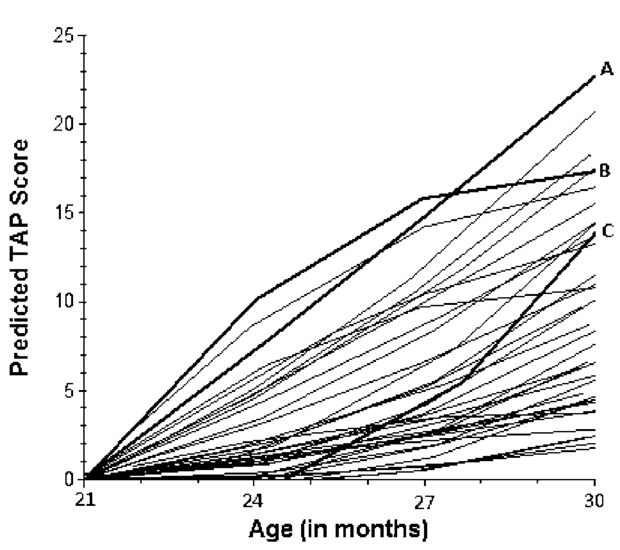
*$p < .05$. **$p < .01$. ***$p < .001$. †$p < .10$.

represents a child who has a slower start and then accelerates over the 9-month span.

Model 3 (Equation 4) examined whether child sex was a significant predictor of between-child differences in the growth parameters. There was no statistically significant effect of sex on the linear growth at 21 months, $b_{11} = –0.04$, $p = .92$, 95% CI [–0.91, 0.82], or on the quadratic growth over time, $b_{21} = –0.02$, $p = .58$, 95% CI [–0.11, 0.06], indicating that sex differences in TAP growth are not apparent at this early age.

A post hoc analysis was conducted to test for differences in the growth parameters between the Rispoli et al. (2009) DeKalb cohort and the current Champaign cohort.

Figure 1. Children's tense agreement productivity (TAP) growth trajectories from 21 to 30 months. Bold lines illustrate different growth patterns. Child A has a strong linear growth rate from 21 to 30 months; Child B shows a strong initial linear growth rate with deceleration; Child C has slow initial linear growth with acceleration.



In a combined no-intercept quadratic model, cohort was added as a predictor at level 2 (see Equation 5). Cohort was coded with 0 = DeKalb and 1 = Champaign, and this variable was grand-mean centered so that the level 2 intercepts, β_{10} and β_{20}, estimate the linear and quadratic effects, respectively, weighted by the proportion of DeKalb to Champaign participants. Therefore, the intercepts are interpreted as the weighted average of the growth parameters across the entire sample. Although it may seem counterintuitive to center dichotomous variables, it can produce more interpretable intercepts (Enders & Tofighi, 2007).

$$\begin{aligned} \mathrm{TAP}_{ti} &= \pi_{1i}(\mathrm{age}_{ti} - 21) + \pi_{2i}(\mathrm{age}_{ti} - 21)^2 + e_{ti} \\ \pi_{1i} &= \beta_{10} + \beta_{11}(cohort)_i + r_{1i} \\ \pi_{2i} &= \beta_{20} + \beta_{21}(cohort)_i + r_{2i} \end{aligned} \tag{5}$$

Table 3 presents a direct comparison of the fixed and random effects for the two cohorts and the combined model. There were no significant differences between cohorts in linear growth at 21 months, $b_{11} = 0.05$, $p = .893$, 95% CI [–0.66, 0.76], nor in overall acceleration between 21 and 30 months, $b_{21} = –0.02$ $p = .535$, 95% CI [–0.02, 0.05]. Thus, the growth estimates demonstrate reliability between the two independent cohorts.

The EB coefficients from Model 2 for the linear and quadratic growth estimates were outputted and used as predictors of the TEGI elicited grammar composite at 36 months to evaluate the validity of the TAP score in predicting grammatical outcomes 6 months later. The EB coefficients from Model 2 instead of Model 3 were used because sex was not a significant predictor of either growth coefficient. These coefficients are the predicted values based on the age of the child and the estimated growth parameters plus the EB residuals. The EB residuals for each individual are the child-level residuals from the TAP multilevel growth model that indicate how each child deviated from the overall predicted linear growth at 21 months and the overall predicted acceleration curve from 21 to 30 months. Therefore, the EB

Table 3. Comparison of growth coefficients (and standard errors) for the two cohorts and the combined model.

Variable	DeKalb Rispoli et al. (2009) cohort (n = 20)		Current study Champaign cohort (n = 37)		Combined model (n = 57)	
Fixed effect						
Linear growth, β_{10}	0.58*	(0.27)	0.64**	(0.22)	0.62***	(0.17)
Cohort					0.05	(0.36)
Quadratic growth, β_{20}	0.07*	(0.03)	0.05*	(0.02)	0.05**	(0.02)
Cohort					–0.02	(0.04)
Random effects						
Linear (VAR), r_1	0.79**		1.46***		1.23***	
Quadratic (VAR), r_2	0.006*		0.01***		0.01***	
Level-1 error, σ^2	5.21		2.52		3.45	
24-month predicted TAP score						
Mean (± 1 *SD*)	2.37	(0–5.03)				
30-month predicted TAP score						
Mean (± 1 *SD*)	10.89	(2.91–18.87)				

Note. Formula for predicted TAP scores: $Y_{it} = \pi_1(\text{age} - 21) + \pi_2(\text{age} - 21)^2$. To provide expectations for variability in TAP scores, the estimated between-child variance component for linear growth was used to estimate ±1 *SD* in the between-child linear growth rates. For purposes of these comparisons, only average rates of acceleration were used. For example, the computation for +1 *SD* at 24 months in the current sample is $Y_{it} = (.64 + 1.208)(24 - 21) + .05(24 - 21)^2 = 1.92 + 0.45 = 5.99$.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

coefficients are the predicted linear slope of the TAP scores at 21 months and the predicted quadratic curvature in the TAP scores from 21 to 30 months for each child.

The linear growth rate at 21 months was negatively correlated with the growth trajectory over the 9-month span from 21 to 30 months, $r(n = 37) = -.84$, $p < .001$. Hence, a child with a negative linear residual and a positive quadratic residual was a child who had a slower than average rate of growth at 21 months but accelerated between 21 months and 30 months faster than average. Conversely, a child with a positive linear residual and negative quadratic residual was a child who had a faster than average rate of growth at 21 months but the child's acceleration was less than average or the child decelerated between 21 and 30 months. Although not all children followed this pattern, the overall negative correlation between linear rate of growth and quadratic growth is evident in Figure 1.

The estimated linear growth on TAP at 21 months was entered first in the regression analysis, and the quadratic growth from 21 to 30 months was subsequently entered as a predictor to determine if acceleration accounted for significant variance in TEGI scores at 36 months, controlling for linear growth. Although sex differences were not apparent in TAP growth between 21 and 30 months, child sex was also included as a predictor in the regression analyses in order to control for the potential influence of sex in early language development.

Initially, child sex was entered into the model and did not account for any significant variance in TEGI scores, $R^2 = .069$, $F(1, 30) = 2.22$, $p = .15$. Subsequently, the linear rate of growth of TAP at 21 months was entered in the prediction model, and it accounted for significant variance in 36-month TEGI scores over and above child sex, $\Delta R^2 = .27$, $F(1, 29) = 11.55$, $p = .002$. Last, the quadratic rate of growth of TAP from 21 to 30 months was added to the model, and this accounted for additional significant variance in 36-month TEGI scores, controlling for both child sex and the linear rate of growth in TAP at 21 months, $\Delta R^2 = .22$, $F(1, 28) = 13.76$, $p = .001$, resulting in a total $R^2 = .553$ or $R^2_{\text{adjusted}} = .506$, $F(3, 28) = 11.57$, $p < .001$. Regression coefficients for the individual variables are detailed in Table 4. This analysis demonstrates that TAP growth as early as 21 months is predictive of TEGI scores obtained 15 months later, with

Table 4. Hierarchical regression of 36-month TEGI scores on TAP growth from 21 to 30 months.

Model	*B*	*SE*	*β*	*t*	*p*
Model 1					
B_0	46.85	6.93			
Sex	14.16[a]	9.51	0.26	1.49	.147
Model 2					
B_0	37.57	6.56			
Sex	15.96	8.20	0.30	1.95	.061
21-month linear growth	12.68[b]	3.73	0.52	3.40	.002
Model 3					
B_0	15.13	8.15			
Sex	17.33	6.84	0.32	2.53	.017
21-month linear growth	30.47	5.72	1.24	5.33	< .001
Quadratic growth	258.57[c]	69.71	0.86	3.71	.001

Note. R^2 model = .55; adjusted R^2 model = .51.

[a]$R^2 = .07$. [b]$\Delta R^2 = .27$. [c]$\Delta R^2 = .22$.

acceleration/deceleration in the individual child growth trajectories from 21 to 30 months contributing additional predictive information.

Discussion

Our growth modeling analyses revealed several developmental findings. First, there is a common developmental trend for the onset of T/A morphemes. Second, there is a gradual increase of T/A morphemes across diverse sentence contexts, with growth accelerating in the third year of life. Finally, the regression analyses demonstrated that the growth of finiteness marking across diverse sentence contexts between 21 and 30 months predicts the accuracy of finiteness marking at age 3 years. In this discussion, we address the implications of these findings for establishing developmental expectations for purposes of assessing and monitoring language growth as well as for evaluating theoretical accounts of early grammatical development.

TAP Growth as a Measure of Early Finiteness: Developmental Implications

The findings of this study, coupled with Rispoli et al. (2009), provide converging evidence for the reliability and validity of TAP growth for assessing the onset of finiteness as children transition out of the one-word stage and begin to produce more diverse sentences. Direct comparison of the two cohorts revealed compelling evidence of replication. In both cohorts, a no-intercept quadratic model best fit the data, with significant linear growth at 21 months and overall acceleration. Significant variability between children in their linear and quadratic coefficients was also present in both cohorts, although greater variability was observed in the current Champaign cohort as compared to the Rispoli et al. DeKalb cohort. Finally, children's individual linear and quadratic growth coefficients were negatively correlated in both cohorts. That is, slower than average linear growth was associated with faster than average acceleration overall, and vice versa.

The increased variability in the Champaign cohort may have been influenced by sample size and methodological differences between the two studies. The Champaign cohort included 37 toddlers, and an examiner joined the parent–toddler dyad after 30 min to create more diverse opportunities for T/A morphemes. In contrast, the DeKalb cohort (Rispoli et al., 2009) included only 20 toddlers, and the entire 1-hr language sample was based on parent–toddler conversational interaction. Yet, the growth coefficients were comparable for the two cohorts, even with a concerted effort to create opportunities for target morpheme use. The larger sample size of the Champaign cohort also resulted in greater precision in estimating the fixed-effects growth coefficients, as indicated by the smaller standard error values. Given the methodological improvement in our current study, we interpret low TAP scores as resulting from the limited strength of the grammatical features of tense and agreement in sentence production rather than limited opportunities to produce T/A morphemes.

The current findings also establish the predictive validity of the type-based analysis for assessing the onset and growth of finiteness with children who are in the earliest period of grammatical development during conversational, play-based interactions. TAP growth coefficients obtained from conversational language samples between 21 and 30 months accounted for 47% of the unique variance in TEGI outcomes, a composite score obtained from standardized probes appropriate for use with preschoolers and school-age children. The proportion of unique variance accounted for was also remarkably similar to Rispoli et al.'s (2009) finding. In the DeKalb cohort, TAP growth coefficients accounted for 45% of the unique variance in 33-month accuracy estimates obtained from conversational language sampling. Together, the findings from the two cohorts provide evidence for the stability of growth in finiteness from the earliest point at which it can be assessed.

The growth modeling results can also be used to establish quantitative expectations for the onset of finiteness, at least for child speakers of Standard American English from college-educated families. For the average child, finiteness can be characterized as emerging between 21 and 24 months, with gradual acceleration to 30 months, corresponding roughly to the period between MLU 1.50 to 3.00. At 24 months of age, the average MLU in parent–child interaction was 1.86, and the predicted TAP score was approximately 2.0 morphemes. By 30 months of age, the average MLU was 2.93, and the predicted TAP score was 9.8 morphemes (see Table 3). More conservative expectations, based on the lower boundary of Model 2's 95% CI, would estimate a predicted TAP score of approximately 1.0 morpheme by 30 months of age.

Criterion-referenced expectations can also be drawn from the raw TAP scores. Across the two cohorts combined, all 57 children (100%) had a productivity score ≥ 1 by 30 months of age, 91% of the children had scores ≥ 3, 84% had scores ≥ 4, and 74% had scores ≥ 5. In other words, the majority of typically developing children (i.e., 74%) can be expected to demonstrate a TAP score of 5 or more by 30 months of age and/or by an MLU of approximately 2.50. Drawing on the psycholinguistic framework and developmental sequence reported in Rispoli et al. (2012), a TAP score of 5 would typically correspond to two or three different uses of copula BE and the first uses of one or more morpheme classes with nonspecific aspect in functional processing (i.e., –3s, *–ed,* auxiliary DO).

These developmental expectations are most appropriate for children from similar family backgrounds. The children in this sample were drawn from homes in which parents were predominantly college educated, and previous research has established a relationship between parents' education levels and greater lexical diversity in their child-directed speech (Hart & Risley, 1995; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). Given that children's sufficiently different uses of T/A morphemes require lexical diversity in grammatical subjects and lexical verbs, the general rate of growth and developmental expectations reported here may be influenced by more lexically diverse input. The children were also required to be monolingual speakers of Standard American English. Future research is needed to determine whether an emphasis on finiteness marking across diverse sentence contexts can be usefully extended to early grammatical assessment with more diverse groups of language learners.

On the other hand, when assessing children from similar demographic backgrounds, clinicians can use the growth-relevant expectations to monitor a child's risk for language impairment, given that pronounced weakness in finiteness relative to MLU expectations is a hallmark that is characteristic of preschoolers with SLI, in particular (cf. Leonard, 1998; Oetting & Hadley, 2009; Rice, 2003). Future research is needed to investigate whether late emergence of finiteness coupled with slower growth thereafter increases the risk for SLI (Hadley & Holt, 2006; Hadley & Short, 2005; Rispoli et al., 2009). Delayed onset of finiteness could be referenced relative to age and/or to MLU by determining whether evidence of finiteness marking has begun to appear (i.e., TAP score of 5) by 30 months or by an MLU of approximately 2.50.

Documenting growth over time is particularly important. In some cases, later than expected emergence may be followed by greater than average growth, as was captured by the negative relationship between the linear and quadratic coefficients in this study. In other cases, late emergence may be followed by average growth. This growth pattern may reflect slow, typical language development. However, clinicians may be more concerned when children present with both significantly later emergence and significantly slower growth. Fuchs and Fuchs (1998) referred to this growth pattern as a *dual discrepancy*. Using the TAP score, Hadley and Holt (2006) observed a dual discrepancy in the growth of finiteness for toddlers with a positive family history of language learning disabilities (i.e., *lower* scores at 30 months and *slower* growth thereafter), after controlling for child sex, receptive language ability, MLU, and maternal education. Thus, later emergence and slower growth in finiteness may be indicative of an increased risk for language impairment.

The stability observed between early TAP growth coefficients and the TEGI outcomes at age 3 years underscores the importance of monitoring growth over time for children with low scores and slow growth who may be at risk for language impairment. In the current study, eight of the 10 children with 30-month predicted TAP scores < 5 had 36-month outcomes below the TEGI cutoff score of 44%, the empirically derived value resulting in the optimal balance of sensitivity and specificity levels for children ages 3;0 (years;months) to 3;5 (i.e., 80% sensitivity, 77% specificity; Rice & Wexler, 2001). Thus, 80% of the children with limited TAP scores at 30 months displayed weaknesses in accurate use of finiteness at age 3 years relative to same-age peers.

Finally, our exploration of child sex as a predictor of early growth in finiteness produced mixed results. The growth model analyses did not indicate that child sex was a significant predictor of TAP growth before 30 months (cf. Model 3); however, when examining the 36-month outcomes, child sex was a significant predictor of TEGI composite scores once both the linear and quadratic growth coefficients were entered in the regression equation. Additional longitudinal studies are needed to understand how child sex interacts with the developmental window under investigation during early childhood. This is particularly important for the period between the ages of 2 and 6 years insofar as otherwise healthy boys are more likely than girls to be identified with late language emergence at age 2 years and to resolve their early language delays by school age (Rice, Taylor, & Zubrick, 2008; Zubrick et al., 2007).

Early Growth in Finiteness: Theoretical Implications

How might our empirical findings be interpreted within current theories explaining the acquisition of tense and agreement? For Wexler (1998), grammatical development is under the control of biological maturation. Our evidence for an age-defined onset and a general growth trajectory for the group is compatible with this maturational perspective. Despite wide variation in vocabulary abilities at 21 months, the children evidenced limited variability in TAP scores. In fact, 73% of the children had a TAP score of 0 at this age, which indicates that the emergence of T/A morphemes is not driven by vocabulary alone. Wexler's explanation for the gradual decline in root infinitives, or sentences lacking T/A morphemes, is complex. On one hand, Wexler posits very early knowledge of inflection (VEKI). That is, Wexler notes that when a child produces a word form indexed by T/A features, such as *goes,* the form is almost always matched to a third person subject. Our data support this claim. Of the 532 verb–3s forms produced by the children in our corpora, 520 (98%) were paired correctly with third person singular subjects. However, if Wexler is correct, and children do have VEKI, then why don't children use inflection in all obligatory contexts? Wexler's (1998, 2003) solution to this paradox is the unique checking constraint (UCC), which weakens with maturation. He posits that the UCC relaxes gradually, so that root infinitives and other associated phenomena do not disappear suddenly, but rather over an extended period of time. In other words, the combination of VEKI and UCC allows the child to have early knowledge, but this knowledge is revealed slowly through a gradual *reduction* of constraint strength.

From the usage-based perspective, developmental growth in T/A morphemes could be interpreted as an increase in the number of rote-learned inflected forms and lexically specific constructions (Ambridge & Lieven, 2011). This alternative hypothesis cannot be ruled out. However, the increasing freedom with which auxiliaries and copula combine with sentence subjects is more difficult to explain unless generalization has already begun, and if so, a usage-based metric for gauging progress in generalization is needed. Usage-based theorists also question whether inflected forms of real words provide sufficient evidence of morphological productivity. For example, Ambridge and Lieven wrote that "studies that use real as opposed to novel items (e.g., Rice & Wexler, 1996) cannot be taken as compelling evidence of morphological productivity" (p. 163). However, even during this period of early growth, some evidence of morphological productivity is apparent in our data. Nineteen overregularizations of *–ed* (e.g., *breaked, comed*) and –3s (e.g., *gots*) appeared in our corpora produced by 15 of the 37 children. The *-ed* overregularizations were produced when the average number of conventional verb-*ed* types was only 1.47 verbs. Thus, overregularizations were observed with only limited production of conventional forms. This weakens the argument that overregularizations are the culmination of a stockpiling of stored rote forms.

Our findings do not resolve the question of whether VEKI and UCC are real, or whether the growth we documented can be reduced to an inventory of rote-learned inflected forms. GML offers an alternative perspective, emphasizing the grammatical features shared across the morphemes in the TAP score within an explicit theory of sentence production (Bock & Levelt, 1994). Grammatical features are embedded in clause structure. In sentence production, lexical forms are not only retrieved, but they must also match the grammatical feature specification of the clause, satisfying the requirements of functional and positional processing (Rispoli & Hadley, 2011; Rispoli et al., 2012). Gradual strengthening of the grammatical features guarantees the correct match between lexical forms and clause structure. We interpret the common developmental trend in the growth of T/A morphemes in diverse sentence contexts as evidence that clausal representations and the grammatical features associated with these representations are increasing in strength in the third year of life. In GML, this process is a form of learning.

Future research should test the unique predictions of GML relative to competing accounts. One unique prediction of GML is *cross-morpheme facilitation*. Because the same grammatical features are shared by different morphemes, the production of one morpheme can, in principle, facilitate the learning of another morpheme. Rispoli et al. (2012) presented evidence that the feature bundle (third person singular present tense) found in copula *is* strengthens the representation of verb–3s. The timing of cross-morpheme facilitation is proximal to the point when sensitivity to the presence of verb–3s in comprehension emerges in children (Sundara, Demuth & Kuhl, 2011), when knowledge of the tense distinction has been demonstrated (Valian, 2006), and when priming between auxiliary BE forms has been observed (Rissman, Legendre, & Landau, 2013). Usage-based accounts do not predict cross-morpheme relationships of this type in the third year of life. GML also introduces a novel hypothesis for how children can benefit from input. Because of cross-morpheme facilitation, input for the highly frequent copula *is* may have an indirect effect on the acquisition of verb–3s (Rispoli & Hadley, in press). Input effects indicate that biological maturation is not solely responsible for the resolution of the optional infinitive stage. Future studies designed to test for input effects and the influence of cross-morpheme facilitation would significantly advance the scientific understanding of the mechanisms underlying morphosyntactic learning.

Conclusion

This study has provided new evidence for the reliability and predictive validity of our conservative, type-based approach to measuring the growth of finiteness in the third year of life. By focusing on the whole suite of T/A morphemes produced across a diversity of sentence frames, a general developmental pattern of early growth in finiteness between 21 and 30 months was revealed, and between-child differences in rates of growth predicted children's accurate marking of finiteness at age 3. As such, this type-based approach can be used to assess young children's early grammatical development.

Acknowledgments

This study was supported by National Study Foundation Grant BCS-082251, awarded to Matthew Rispoli and Pamela A. Hadley. Portions of this article were presented at the 2012 Symposium for Research in Child Language Disorders, Madison, WI, and at the 2012 American Speech-Language-Hearing Association's Annual Convention, Atlanta, GA. We are grateful for the commitment of the participating parents and children to this project and to the numerous students who assisted with data collection and transcription.

References

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, United Kingdom: Cambridge University Press.

Bauer, D., Goldfield, B., & Reznick, S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics, 23,* 313–335.

Bleile, K. (1995). *Manual of articulation and phonological disorders*. San Diego, CA: Singular.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89,* 1–47.

Bock, J. K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gersenbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Bornstein, M., Hahn, C., & Haynes, O. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24,* 267–304.

Bricker, D., & Squires, J. (1999). *Ages & Stages Questionnaires: A parent-completed, child-monitoring system* (2nd ed.). Baltimore, MD: Brookes.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82,* 711–733.

Caprin, C., & Guasti, M. (2009). The acquisition of morphosyntax in Italian: A cross-sectional study. *Applied Psycholinguistics, 30,* 23–52.

Cleave, P., & Fey, M. (1997). Two approaches to the facilitation of grammar in children with language impairments. *American Journal of Speech-Language Pathology, 6,* 22–32.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63,* 552–543.

Eisenberg, S., Guo, L., & Germezia, M. (2012). How grammatical are 3-year-olds? *Language, Speech, and Hearing Services in Schools, 43,* 36–52.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories* (2nd ed.). Baltimore, MD: Brookes.

Fuchs, L., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13,* 204–219.

Gladfelter, A., & Leonard, L. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of Speech, Language, and Hearing Research, 56,* 542–552.

Guo, L., Van Horne, A., & Tomblin, B. (2011). The role of developmental levels in examining the effect of subject types on the production of auxiliary is in young English-speaking children. *Journal of Speech, Language, and Hearing Research, 54,* 1658–1666.

Hadley, P., & Holt, J. (2006). Individual differences in the onset of tense marking: A growth curve analysis. *Journal of Speech, Language, and Hearing Research, 49,* 984–1000.

Hadley, P., Rispoli, M., Fitzgerald, C., & Bahnsen, A. (2011). Predictors of morphosyntactic growth in typically developing toddlers: Contributions of parent input and child sex. *Journal of Speech, Language, and Hearing Research, 54,* 549–566.

Hadley, P., & Short, H. (2005). The onset of tense marking in children at risk for SLI. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.

Holt, J. K. (2008). Modeling growth using multilevel and alternative approaches. In A. A. O'Connell & D. B. McCoach (Eds.), *Quantitative methods in education and the behavioral sciences* [Issues, Research and Teaching Series: Vol. 3, Multilevel Analysis of Educational Data] (pp. 111–159). Charlotte, NC: Information Age.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61,* 343–365.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69.

Lapointe, S., & Dell, G. (1989). A synthesis of some recent work in sentence production. In G. Carlson & M. Tannehaus (Eds.), *Linguistic structure in language processing* (pp. 107–156). Dordrecht, the Netherlands: Kluwer.

Leonard, L. (1998). *Specific language impairment in children*. Cambridge, MA: MIT Press.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin, Germany: Mouton de Gruyter.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43*(1–2, Serial No. 174).

MacWhinney, B. (2000). *The CHILDES project* (3rd ed.). Mahwah, NJ: Erlbaum.

Miller, J., & Iglesias, A. (2010). Systematic Analysis of Language Transcripts (Research Version 2010) [Computer software]. Middleton, WI: SALT Software.

Oetting, J., & Hadley, P. (2009). Morphosyntax in child language disorders. In R. G. Schwartz (Ed.), *The handbook of child language disorders* (pp. 341–364). New York, NY: Psychological Press.

Pine, J., Conti-Ramsden, G., Joseph, K., Lieven, E., & Serratrice, L. (2008). Tense over time: Testing the agreement/tense omission model as an account of the pattern of tense-marking provision in early child English. *Journal of Child Language, 35,* 55–75.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S., & Jackendoff, R. (2009). The reality of a universal language faculty. *Behavioral and Brain Sciences, 32,* 465–466.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford, United Kingdom: Blackwell.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2011a). HLM 7: Hierarchical linear and nonlinear modeling [Computer software]. Lincolnwood, IL: Scientific Software International.

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2011b). HLM for Windows (Version 7.0) [Computer software]. Lincolnwood, IL: Scientific Software International.

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research, 40,* 556–566.

Rice, M. (2003). A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 63–95). Mahwah, NJ: Erlbaum.

Rice, M., Taylor, K., & Zubrick, S. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 50,* 1562–1592.

Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English speaking children. *Journal of Speech and Hearing Research, 39,* 1239–1257.

Rice, M., & Wexler, K. (2001). *The Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: The Psychological Corporation.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431.

Rispoli, M. (1994). Structural dependency and the acquisition of grammatical relations. In Y. Levy (Ed.), *Other children, other languages: Issues in the theory of language acquisition* (pp. 265–301). Hillsdale, NJ: Erlbaum.

Rispoli, M. (2005). When children reach beyond their grasp: Why some children make pronoun case errors and others don't. *Journal of Child Language, 32,* 93–116.

Rispoli, M., & Hadley, P. (2011). Toward a theory of gradual morphosyntactic learning. In I. Arnon & E. Clark (Eds.), *Experience, variation and generalization: Learning a first language* (pp. 15–33). Amsterdam, the Netherlands: John Benjamins.

Rispoli, M., & Hadley, P. (in press). Input effects on the acquisition of finiteness. In C. Chu, C. Coughlin, B. L. Prego, U. Minai, & A. Tremblay (Eds.), *Proceedings of the 5th Generative Approaches to Language Acquisition North America* (pp. xx–xx). Somerville, MA: Cascadilla Press.

Rispoli, M., Hadley, P., & Holt, J. (2009). The growth of tense productivity. *Journal of Speech, Language, and Hearing Research, 52,* 930–944.

Rispoli, M., Hadley, P., & Holt, J. (2012). Sequence and system in the development of tense and agreement. *Journal of Speech, Language, and Hearing Research, 55,* 1007–1021.

Rissman, L., Legendre, G., & Landau, B. (2013). Abstract morphosyntax in two-and three-year-old children: Evidence from priming. *Language Learning and Development, 9,* 278–292.

Shriberg, L., Kwiatkowski, J., & Hoffmann, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research, 27,* 456–465.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1157–1256). Hillsdale, NJ: Erlbaum.

Sundara, M., Demuth, K., & Kuhl, P. K. (2011). Sentence-position effects on children's perception and production of English third person singular *–s. Journal of Speech, Language, and Hearing Research, 54,* 55–71.

Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16,* 247–277.

Theakston, A., & Rowland, C. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study: Part 1. Auxiliary BE. *Journal of Speech, Language, and Hearing Research, 52,* 1449–1470.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Valian, V. (2006). Young children's understanding of present and past tense. *Language Learning and Development, 2,* 251–276.

Van Valin, R. (2009). Some remarks on universal grammar. In J. Guo, E. Lievan, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language* (pp. 311–320). New York, NY: Taylor & Francis.

Wexler, K. (1994). Optional infinitives, head movement, and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 305–350). Cambridge, United Kingdom: Cambridge University Press.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106,* 23–79.

Wexler, K. (2003). Lenneberg's dream: Learning, normal language development and specific language impairment. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Towards a definition of specific language impairment* (pp. 11–61). Mahwah, NJ: Erlbaum.

Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language, 30,* 75–115.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford, United Kingdom: Oxford University Press.

Yang, C. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Sciences, 8,* 451–456.

Zubrick, S., Taylor, K., Rice, M., & Slegers, D. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 51,* 394–407.