Individual differences in syntactic ability and construction-learning: an exploration of the relationship

Nick Riches
nick.riches@newcastle.ac.uk

Laura Jackson
laura.jackson@yahoo.co.uk

Department of Speech and Language Sciences
School of Education Communication and Language Sciences
Newcastle University
United Kingdom
NE1 7RU
Individual differences in syntactic ability and construction-learning: an exploration of the relationship

An important test for any linguistic theory is whether it can account for variation in language abilities. Such variation is large, and evident from an early age. Around one in ten young children exhibit severely delayed expressive language (Fenson et al., 1993). Though many of these ‘late talkers’ appear to catch up with peers on standardised assessments, language skills remain weak into adolescence and beyond (Rescorla & Turner, 2015). About 7% of children experience extreme language difficulties necessitating early intervention and continued support (Tomblin, Records, Buckwalter, & Zhang, 1997). The label ‘Specific Language Impairment’ (SLI) is often applied to these children. Their spoken language is often strikingly ungrammatical, containing morphosyntactic errors such as omission of past tense –ed. Though such errors decline with age, poor performance on standardised assessments persists into adolescence (Durkin & Conti-Ramsden, 2007).

Though it is tempting to view language abilities in dichotomous terms, e.g. language-impaired versus language-typical, variation is nonetheless observed in non-clinical populations. Adults with low educational attainment struggle to comprehend complex structures such as passives (Dabrowska & Street, 2006). Even university undergraduate students, presumably with strong language skills, misinterpret passives describing implausible events (Ferreira, 2003). This suggests a reliance on heuristics, e.g. animacy and event probability, rather than detailed structural parsing. These studies question a dichotomous view of language variation, suggesting a single continuous dimension of language ability. Moreover, if we view linguistic competence as the ability to access abstract structural representations, rather than low-level cues, the above studies suggest that even competent native speakers do not attain the ‘end state’.
Though individual variation in language abilities is well-documented, there is still considerable disagreement regarding its causes. Broadly speaking, researchers have proposed either limited knowledge (competence), or deficient processing (performance). This dual perspective reflects a long tradition in psycholinguistics whereby competence and performance are viewed as distinct. Evidence for this dichotomy is provided by sentences which, though well-formed, are difficult to process, e.g. *the rat the cat the dog chased killed ate the malt* (Miller & Chomsky, 1963). If we use brackets to highlight its structure, e.g. *the rat [the cat [the dog chased] killed] ate the malt*, we can see that the sentence is grammatically well-formed, consisting of multiple centre-embedded object relative clauses. However, if the structure is not highlighted, the sentence is extremely difficult to process. This type of sentence suggests a separation between competence, which enables us to judge it as grammatical, and performance, which is overwhelmed by the multiple dependent relationships between subjects and verbs. The competence / performance dichotomy pervades research into individual differences. For example, it has been suggested that children with SLI fail to mark tense because, until a certain age, they do not know that it is obligatory (Rice & Wexler, 1996). However, the same phenomenon can be explained by assuming that these children have difficulties processing morphemes involving rapid articulatory movements (Leonard & Eyer, 1996).

Though the competence-performance dichotomy has been used to frame explanatory theories of language difficulty, many researchers are sceptical of this distinction. Usage-based accounts, which argue against the existence of Universal Grammar, are particularly opposed to this categorical approach (Ibbotson, 2013). They argue that usage-related factors such as input frequency play a major role in shaping the language system, and therefore usage and competence are essentially indivisible. Other models have proposed that language structure is strongly influenced by processing constraints such as working memory or parallel
processing (Jackendoff, 2007; O’Grady, 2005) thereby questioning a strict competence-performance dichotomy.

Though the distinction between competence and performance has been repeatedly questioned, models which conflate these two phenomena face a paradox. If competence and performance are indeed ‘two sides of the same coin’ how do we explain individual variation? If the two phenomena are perfectly correlated, how can one vary without the other? One solution is to propose that differences in language ability reflect differences in language experience. In support of this, the ability to process complex structures such as relative clauses can be boosted by increased exposure (Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Likewise, poor understanding of complex syntax by adults with low educational attainment may reflect variation in print exposure (Dabrowska & Street, 2006). However, variation in experience cannot explain most cases of SLI, as there is little evidence for impoverished input (Leonard, 2014).

An alternative possibility is that individual variation reflects variation in language-learning ability. This claim will be referred to as the ‘Learning Hypothesis’. There is now substantial research indicating that language acquisition is supported by powerful implicit learning mechanisms, and moreover, these vary substantially across individuals. In the lexical domain, fast-mapping studies indicate a mechanism which rapidly acquires form-meaning pairs, possibly after a single exposure (Carey & Bartlett, 1978). These mappings are temporally-robust with information retained over intervals ranging from 24 hours to a week (Carey & Bartlett, 1978; Woodward, Markman, & Fitzsimmons, 1994). In the syntactic domain, implicit learning mechanisms have been revealed by structural priming studies (Bock, 1986). Many have suggested that structural priming effects, whereby participants recycle recently-encountered structures, reflect implicit learning (e.g. Bock & Griffin, 2000). Like lexical fast-mapping, priming effects are instantaneous (observable after a single item),
and long-lasting, with priming observed after ten intervening items (Bock & Griffin, 2000). In view of these similarities, some researchers have begun to use the term ‘fast-mapping’ to refer to syntactic learning (Casenhiser & Goldberg, 2005; Kim, O’Grady, Deen, & Kim, 2017). This use of terminology implies that lexical and syntactic learning are underpinned by similar processes.

There is converging evidence that implicit learning mechanisms are linked to individual variation in language abilities. Cumulative structural priming effects, i.e. the rate at which priming increases over successive presentations of the structure, strongly predict syntactic comprehension in language-typical children (Kidd, 2012). Performance on implicit artificial grammar learning tasks by adults predicts comprehension of complex structures (object relatives) (Misyak & Christiansen, 2012). Moreover, implicit learning mechanisms seem to be weaker in language-impaired individuals. Adolescents with language difficulties are poor at acquiring artificial grammars (Hsu, Tomblin, & Christiansen, 2014), and language-impaired children perform poorly on statistical learning tasks involving lexical segmentation (Evans, Saffran, & Robe-Torres, 2009). In the lexical domain, children with SLI perform poorly at lexical fast-mapping (Rice, Buhr, & Nemeth, 1990).

This study aims to test the Learning Hypothesis by investigating the relationship between construction-learning abilities, and performance on two standardised assessments of syntactic abilities: the Test of Reception of Grammar (TROG: Bishop, 2003), which assesses sentence comprehension, and the Renfrew Action Picture Task (RAPT: Renfrew, 1997), which assesses sentence production. These tests investigate ‘static’ language skills: the ability to produce or comprehend a particular structure at a particular moment in time. According to the Learning Hypothesis, the ability to acquire a construction should be strongly predictive of static assessments of syntactic abilities. While many studies have investigated grammatical learning in laboratory settings, e.g. the AGL studies cited above, and studies investigating
fast-mapping of constructions (e.g. Casenhiser & Goldberg, 2005; Wonnacott, Boyd, Thomson, & Goldberg, 2012) they have not focused on individual variation, or whether learning is associated with performance on static tasks. This study will address this gap in the literature.

A further aim is to enhance ecological validity. Previous studies have taught artificial structures consisting of nonwords (AGL tasks), and invented grammatical constructions (e.g. Wonnacott et al., 2012). By maximising experimental control, these sacrifice ecological validity. For example, Wonnacott et al. (2012) taught the novel construction *Verb NP1 NP2* with the meaning that NP2 approaches NP1 in the manner of the verb. However, this can be paraphrased using the English locative, e.g. *the cat crawled towards the mouse*. Synonymy with an existing construction is problematic as language obeys the principle of contrast, whereby different forms signal different meanings (Clark, 1987). In this sense the construction *Verb NP1 NP2* is un-language-like.

Another source of artificiality is the relationship between formal and functional idiosyncrasy. Many constructions which ‘bend’ grammatical rules have idiosyncratic meanings. For example, *he’s not the boss of me*, has a very specific meaning (challenging authority) compared to the syntactically more conventional *he’s not my boss*. The ditransitive is also unconventional in placing the indirect object immediately after the verb, e.g. *give me it*. It expresses a wider range of meanings than the more syntactically-conventional prepositional dative (*give it to me*), e.g. negated transfer (*she denied me entry*), or potential transfer (*she promised me a massage*) (A. E. Goldberg, 1995).

The language-likeness of the stimuli is likely to impact on the learning process. There is evidence that children map novel constructions onto known constructions. For example, children may acquire the passive construction by mapping it to already known constructions which exhibit partial formal and functional overlap (Abbot-Smith & Behrens, 2006). This
mapping is dependent on the cognitive process of analogy (Gentner, 1983). Likewise, children may find the prepositional dative easier to comprehend than the ditransitive because it exhibits formal overlap with transitives (Rowland & Noble, 2010). However, a syntactically novel construction such as *Verb NP1 NP2* cannot be easily assimilated into existing constructional networks. An alternative route, consistent with the principle of contrast, is to assume that the novel form signals a novel meaning. However, this is also unavailable, as the novel form is mapped to a conventionalised meaning. Given that neither of these options are available, children are likely to be dependent on learning mechanisms which are not typically employed during language acquisition. For example, the construction may be learnt by explicitly mapping each element of the construction onto the pre-existing transitive construction (*NP1 Verb NP2*) which in turn is mapped to a conventionalised meaning. Such a process may depend on metalinguistic awareness and short-term memory to a much greater extent than the analogical processes typically employed in language learning.

Controlled studies with artificial stimuli are vital for improving our understanding of learning mechanisms. However, it is extremely difficult to create novel stimuli which are fully language-like. To overcome this difficulty, this study taught low-frequency constructions which are assumed to be unfamiliar to young children. Though this sacrifices control, it nonetheless boosts the ecological validity of the learning task, which enhances sensitivity to implicit learning mechanisms. Though unfamiliar, as opposed to novel items have been used in studies of word-learning (Rice et al., 1990), this is, we believe, the first time that such an approach has been adopted by a study of study of syntactic learning.

In summary, the study asks the following research questions

- To what extent is there individual variation in construction-learning abilities?
- To what extent is this variation associated with performance on static syntactic assessments?
Can we use unfamiliar constructions to investigate construction-learning?

The final research question is closely linked to the first two. Firstly, wide individual variation in construction-learning would provide evidence that the task taps a cognitive mechanism which is also subject to individual variation. Such variation is predicted by the Learning Hypothesis. Secondly, a strong association with other language assessments supports the concurrent validity of the task.

Further exploratory analyses investigated cognitive mechanisms involved in construction-learning. The two chosen constructions (see below), differ in terms of whether they contain a non-adjacent relationship, whereby two closed slots are separated by an open slot. This type of configuration is widely observed in natural languages. For example, in present progressive sentences, e.g. *He is [eat]-ing*, *is* and *–ing* are closed class elements separated by an open verb slot. The learning of artificial grammars involving non-adjacent relationships is closely associated with overall syntactic abilities (Hsu et al., 2014; Misyak & Christiansen, 2012), suggesting that the ability to acquire such relationships is a fundamental skill supporting language-learning. It was therefore hypothesised that the ability to acquire the non-adjacent construction would be more strongly correlated with static assessments of syntactic abilities.

**Procedure**

**Learning task**

As argued above, the use of invented constructions may undermine the ecological validity of the learning task, and hence its sensitivity to underlying mechanisms. To overcome this, we taught two real English constructions likely to be unfamiliar to young children, the ‘what makes’ and the ‘just because’ constructions:

1. (1) **What makes you** [ think [ you can run fast ] ] ?

2. (2) [ Just because my legs are short ] doesn’t mean [ I can’t run fast ].
Underlining is used to highlight lexically-specific chunks (i.e. remaining constant across exemplars), while the brackets show embedded clauses. To investigate variability within these chunks frequency counts were derived from the Corpus of Contemporary American English (Davies, 2008) containing approximately 500 million words. There were 624 instances of the what makes construction, and 254 instances of the just because... doesn’t mean... construction. For the latter search, the search string specified 1 – 4 words in the first open slot, the maximum allowable by the online search engine. 16 of the what makes sentences (2.5%), contained an alternative pronoun (him/her). Substitutions of the verb: what leads you to think..., Just because... doesn’t imply... were not attested. These data suggest that the underlined items demonstrate little variation, and can therefore be regarded as lexically-specific chunks.

Due to the length and complexity of the constructions, they are unlikely to be used in either child or child-directed speech. For example, both involve multiple embedded clauses. The just because construction involves embedding in subject position, which is rare in English. Furthermore, it contains negative raising, whereby the negative operator takes wide scope (It is NOT the case that just because my legs are short means I can’t run faster than you). Both constructions have a complex pragmatic function: to challenge another person’s assumptions, e.g. that the hearer can run faster than the speaker. Additionally, they are used to critique the interlocutor’s reasoning / argumentation. Though young children can engage in argumentative discourse, the ability to critique another person’s argument emerges during adolescence (Felton, 2004). Consequently, the target constructions have a discourse function far in advance of young children’s pragmatic abilities.

As mentioned, above, the constructions differ in terms of whether they contain non-adjacent chunks. While the just because construction inserts a clause between just because
and doesn’t mean, the what makes you think construction consists of a single chunk which is followed by a clause.

To verify the intuition that the target constructions are rare in child language, a search was conducted on the entire set of British English and American English corpora from the CHILDES database (MacWhinney, 2000, data downloaded 28th March, 2016). This consists of 856,613 child utterances (lines beginning *CH, identified with the UNIX grep command), out of a total of 2,534,302 utterances altogether (lines beginning *). Though many of the children in the CHILDES database are younger than the children in the study it represents the best means of ascertaining the frequency of these constructions in child-directed speech. The UNIX grep command identified target utterances using the search strings ‘what makes you think’ and ‘just because’. There were 21 instances of ‘what makes you think…?’ questions, all of which were produced by adult interlocutors. This amounts to 0.001% of adult utterances. There were 10 instances of ‘just because… doesn’t mean…’, which again were produced by adult interlocutors, amounting to 0.0006% of adult utterances. Neither utterance was produced by a child. Given these statistics, it is reasonable to assume that young children rarely hear the target constructions, and almost never produce them.

From a theoretical perspective, the chosen structures involve idiosyncratic form-function mappings as they have specific pragmatic functions which are not wholly predictable from their form. For example (1) does not function like a typical question as you would not supply the questioned NP by responding ‘my athletic ability’. Such idiosyncrasy is important from the perspective of construction-based accounts of language-learning (e.g. Ellis, 2003; Adele E. Goldberg, 2006) (e.g. Goldberg, 2006) which argue that learning a language involves learning constructions; form-function mappings whereby ‘some aspect of its form or function is not strictly predictable from its component parts or from other
constructions’ Goldberg, 2003, p. 219). Consequently, these can be regarded as genuine constructions which tap basic construction-learning mechanisms.

**Materials**

A picture book called “Little Hippo’s Big Adventure” was created (see online materials in appendix). This told the story of a little hippo who is ridiculed by all the other animals for being short and unathletic. However, the arrival of the animal Olympics gives him a chance to prove himself. He competes in numerous events and wins. There are six events, and six different competitors. Within each event both constructions were modelled. The story was repeated to allow sufficient observations to assess construction-learning ability. Thus each child heard each structure twelve times, and was required to produce each structure on twelve occasions (24 elicitations altogether).

The story is designed to elicit the target constructions. For example, the first event is the high jump, and the hippo competes against the kangaroo. The kangaroo says ‘What makes you think you can jump high? Your legs are too short.’ The hippo replies ‘Just because my legs are short doesn’t mean I can’t jump high.’ Other events are swimming, gymnastics, running, the javelin, and tennis. Online appendices provide further information about the story episodes and prompts, the scoring protocol, storybook pictures and the script followed by the experimenter.

The story was designed to elicit the target constructions in a naturalistic context which illustrates their meaning. For example, before the characters produce the ‘what makes you think’ construction, the child is told that they are ‘mean and horrible’. This emphasises the confrontational pragmatic function of the construction. Likewise, the children are told that the hippo is angry, and this is also shown in the picture.
Participants

49 children were recruited from Year 1 classes in three separate schools in the North East of England. The invitation letters to the school and parents specified that we were looking for children with age-appropriate language skills who did not have learning difficulties. The mean age of the participants was 5;2 with a standard deviation of 3.25 months. The youngest child was 4;8, and the oldest child was 5;8.

Procedure

The procedure was modelled on Dynamic Assessment (DA) (Lidz, 1991). The term covers a range of methodologies used to assess children’s learning abilities. The current study adopted a specific DA procedure called graduated prompting (Gutierrez-Clellen & Peña, 2001). This involves adjusting the amount of support given to the child based on their performance. This approach is Vygotskyan in that it attempts to gauge the child’s Zone of Proximal Development: what they are capable of achieving with maximum scaffolding (Vygotsky, 1978).

The children were shown the story book and were told they were going to read a story called ‘Little Hippo’s Big Adventure’. At periodic intervals the children were required to repeat the target construction (elicited imitation). At first glance, this technique appears to involve Short-Term Memory (STM), a memory system dedicated to maintaining recently-encountered information in a temporarily activated state, with little to no involvement of Long-Term Memory (LTM) (Atkinson & Shiffrin, 1968). However, there is substantial evidence to demonstrate that it taps underlying linguistic knowledge. For example, in a series of experimental studies, Potter and Lombardi proposed that, because STM is too limited to support the recall of sentences above a certain length, repetition involves reconstruction of the sentence from representations in long-term memory (LTM) (Lombardi & Potter, 1992;
Potter & Lombardi, 1990, 1998). Studies of second language learners also indicate a strong role for representations in LTM. Elicited imitation performance is strongly associated with implicit knowledge, assessed by tasks such as a speeded grammaticality judgement (Spada, Shiu, & Tomita, 2015). Finally, the syntactic errors children make during repetition also suggest the involvement of LTM, e.g. *there’s the cat that dog chased* --> *there’s the dog that chased the cat*, which transforms a non-canonical object relative into a canonical subject relative with no lexical changes (e.g. Frizelle & Fletcher, 2014; Riches, 2012). Such syntactically-driven patterns would not occur if sentences were merely being parroted from representations in STM. Given this evidence, improvements in repetition over time were assumed to reflect learning of the target constructions. Given that the adopted measure of learning is potentially confounded with STM, analyses were conducted to verify that construction-learning and STM were separable factors. These are detailed in the final part of the results section.

The story began with a picture which was designed to teach the elicited imitation methodology. Each time a repetition was elicited, the experimenter pointed at the animal who had just spoken. This procedure was used to indicate that it was the child’s turn to repeat the sentence. The experimenter only proceeded to the actual story when the children demonstrated a good understanding of the procedure, i.e. they were able to repeat the target utterance after the prompt.

An outline script (Appendix 4) was generated for the experimenter. However, within this script there was room for more spontaneous interaction with the child. For example, the children were asked to guess which animal competed in which event, and who the winner was. Thus the story was kept consistently engaging for the child.

As mentioned above, the story was repeated to provide sufficient observations to assess the child’s learning trajectory. To implement graduated prompting the narrative was
divided into two blocks of three events. As the story was repeated there were four blocks altogether. The level of support the child received varied according to their performance within the block. If, at the end of the block, the child had produced two perfect sentences on the past three attempts they ‘stepped up’ to the higher level, which involved reduced support. If the child had failed to produce a nearly perfect sentence (≤1 word-level errors) on the past three elicitations, they ‘stepped down’ to the lower level, which involved greater support. If children’s performance fell within the step-up and step-down criteria they stayed at the current level. These criteria were determined during piloting by trial and error, and were fine-tuned to maximally differentiate those children who required a lot of support from those who did not. See appendix 5 for a summary of these criteria.

The children started at the Elicited Imitation level, where they repeated what the characters had said. If they performed well at the first block they stepped up to the highest level; Elicitation. This consisted of three stages. Firstly the child was asked an open-ended question without a prior model; *What did the kangaroo say?* If the child could not produce a near perfect response (≤1 word-level errors), the question was repeated and they were prompted using the first two words of the target utterance. Out of these two responses, the one with the lowest error rate was selected for scoring, bearing in mind that children were not credited with the two-word prompt. Finally, however well the child had performed, the experimenter produced the full utterance. This was to ensure that the child always had a model to learn from.

If the child performed below-criterion at Elicitation Imitation, they moved to the next level down; Imitation plus Gesture. This was a straightforward Elicited Imitation task, but additional support was provided by breaking the utterance up into constituent clauses, inserting a break between the clauses, stressing the first word of the clause, and manually signalling stress by lightly thumping the ground / table with their fist
**What** makes you think *(pause) you* can jump higher than me?

**Just** because my legs are short *(pause)* doesn’t mean I can’t run fast

The maximally-supportive level was Imitation plus Prompt. At this level, the sentence was presented, and during the repetition attempt the experimenter supplied the first two words *(what makes / just because)* for which credit was given during scoring, i.e. responses were scored as if the children had produced these words. In addition, at this level, the gesture, intonation and pause prompts used during the previous level were also employed.

*Scoring*

The child’s distance from the target was measured using the Levenshtein Distance in Words (LDw) (Levenshtein, 1966; Riches, 2012). This algorithm automatically counts how many word additions omissions and substitutions are required to transform one sentence into another. For level four (question), the model sentence was provided after the child’s attempt to answer the question, and therefore the child could produce a sentence which was grammatically correct, and used the target structure, but which differed from the model, e.g. *just because my tail is small doesn't mean I can't swim fast* → *just because my tail is much too small doesn't mean I'm not good at swimming*. A protocol was devised to award the children a perfect score (no errors) for such sentences. However, no child was entirely successful in producing either of the structures during the elicitation stage. Therefore all sentences were scored for accuracy in comparison with the model sentence. It should be noted that this is a fairly strict scoring scheme as lexical departures from the model are penalised even if they are consistent with the situation described in the story.

A further measure of Elicitation Performance was devised to encapsulate both the level of support and the child’s repetition accuracy in a single measure. This was calculated using the following formula:
The second part of this formula is repetition accuracy, which approximately corresponds to the proportion of the sentence correctly recalled. This metric creates a scale between 1 and 4, with the integers (1, 2, 3 & 4) corresponding to the level of support and the numbers after the decimal place corresponding to the child’s performance within that level. For example, if the value of an observation is 1.3, the child is being given maximal support (level 1 = prompted elicitation), and is correctly producing about a third of the sentence. If the value of an observation is 3.7, the child is at level 3 (Elicited Imitation) and is correctly producing about 70% of the sentence. This measure is unusual in combining two different constructs: the amount of support a child requires to produce the construction, and the accuracy of their responses. Given that both of these constructs may reflect children’s knowledge of the target constructions, a measure combining the two was assumed to possess greater sensitivity. Finally, Construction Learning Ability itself was quantified as the mean Elicitation Performance in the final block averaged over both constructions. This was based on 6 data points; 2 constructions x 3 episodes. The Results section will refer to experimental variables with initial capitals, e.g. Construction-Learning Ability, while the underlying construct will be referred to with initial lower-case letters, e.g. learning ability.

Assessments of comprehension and production

The Test of Reception of Grammar (TROG – written form), (Bishop, 2003) was used as a measure of grammatical comprehension. The experimenter produces a sentence and the child must choose the correct picture from a choice of four. Distractors are designed to elicit common comprehension errors. For example, for the sentence the man looks at the horse and is running, one of the pictures shows the horse running, a common error type which involves linking the verb to the closest preceding Noun Phrase. The test has been standardised on a
large sample (n = 894 for entire age range, n = 110 for range 5;0 to 5;11), and yields an internal consistency (Cronbach’s alpha) of 0.88. The test presents numerous structures and syntactic items including spatial prepositions (e.g. the cup is in the box), transitives (e.g. the girl pushes the box), subject relative clauses (e.g. the horse chasing the girl is big), not only X but also Y sentences (e.g. the pen is not only small but also blue), comparatives (e.g. the pencil is longer than the knife), passives (e.g. the cow is chased by the girl), coordinated clauses (e.g. the boy is chasing the dog, and is jumping), pronoun forms (e.g. they are carrying him), complement clauses (e.g. the boy sees that the elephant is touching him) and object relative clauses (e.g. the man the elephant sees is eating), in approximate order of complexity. Given that it includes such a wide range of structures, it therefore provides a comprehensive measure of receptive grammatical abilities.

The Renfrew Action Picture Test (Renfrew, 1997) was used as a production test. It has been standardised on 594 children (n = 117 for age range 5;0 – 5;11). Though not often used as a screening measure, it is nonetheless widely used by clinicians to provide a snapshot of children’s expressive abilities. The test consists of showing a series of ten cards to the participants and asking them what is happening, with the picture forming the only prompt. The pictures are designed so that target items are likely to occur in the response. For example one picture shows a woman lifting a boy so that he can put a letter in a letter box. This elicits a clause expressing purpose, e.g. so that he can put the letter..., or to put the letter.... . The items elicited range from morphemes (progressive –ing, plural –s and irregular past participle -en) to complex constructions such as passives, and relative clauses. It therefore covers a wide range of structures like the TROG. It has a complex scoring algorithm where points are awarded for each item, and credit is also given for partially correct responses. Due to its ecological validity, wide range of structures and sensitive scoring algorithm it can be regarded as a good assessment of general expressive abilities. However, it should be noted
that standardisation is not optimal, e.g. standard scores are calculated for 6 month bins, and reliability measures are not provided.

Both the TROG and RAPT are commercially sensitive and therefore items cannot be listed in full.

Reliability measures

A second rater transcribed the data for 6 children, amounting to 12.7% of child utterances. This process was blinded such that the second rater did not look at the first rater’s transcriptions. Differences arose on 32% of experimental turns. However, differences were minor, with a difference rate of 0.52 words per sentence, calculated by using the LDw comparing the experimenter and interrater’s scores. This amounts to a word-level agreement rate of 94% \( ((1 - \text{mean difference-in-words per sentence} / \text{mean number of words per child response}) \times 100) \). It was not possible to calculate a measure such as Cohen’s Kappa because there is no baseline probability reflecting a random response, e.g. the probability of choosing a correct item from a choice of two. Elicitation Performance was calculated automatically from the transcription using the Levenshtein Distance-in-words, therefore reliability is assumed to be optimal.

For each child, the second rater also checked whether the step-up / step-down rules had been adhered to. This yielded an agreement rate of 88.6% (Cohen’s Kappa = 0.84 based on a chance agreement of 0.33, corresponding to the three possible outcomes: step up, step down, or remain).

Results

Overview

A series of analyses was conducted. Firstly, within-group variation in Construction-Learning Ability was investigated. Subsequent analyses addressed the correlation between
Construction-Learning Ability and performance on the comprehension and production tasks. Regarding interpretation of effect sizes, possibly the most widely used criteria are those of Cohen (1988) (‘small’; $r = 0.1$, $R^2 = 0.01$, ‘medium’; $r = 0.3$, $R^2 = 0.09$, ‘large’; $r = 0.5$, $R^2 = 0.25$). The same guidelines apply to the ‘rho’ statistics produced by Spearman’s correlations. As these values are arbitrary, meta-analytic study of effects sizes in second language research (Plonsky & Oswald, 2014) suggested using the following; “small” $r = 0.25$, “medium” $r = 0.40$, and “large” $r = 0.60$. These correspond roughly to percentiles, with studies at the 25th percentile finding an effect size of 0.25, studies at the 50th percentile finding an effect size of $r = 0.40$, and studies at the 75th percentile finding an effect size of $r = 0.60$. In addition to the correlations, further analyses addressed individual variation in Construction-Learning Ability, and variation in learning rate as a function of structure. Finally, variation in performance across the structures was investigated.

*Individual differences in Construction-Learning Ability*

Figure 1 shows the learning trajectories of the children in the study. The lines show mean Elicitation Performance per story episode. Participants are grouped by quintiles based on Construction-Learning Ability, and group size and mean ages are shown for each quintile. This visual method demonstrates variation in construction-learning ability, though the use of quintiles, as opposed to other units such as terciles, deciles etc., is arbitrary. Variation is wide, with children in the top quintile proceeding rapidly to the least supportive condition, while children in the bottom quintile demonstrate little learning, even in the most supportive condition. There appears to be little effect of age on learning rate.

Figure 2 shows Construction-Learning Ability by construction in the form of histogram. While the *what makes you think* construction demonstrates a bimodal distribution, Construction-Learning Ability for the *just because* construction presents with a more haphazard pattern.
**Association between Construction-Learning Ability, comprehension and production**

Partial correlations controlling for the effect of age-in-months are presented in Table 1. Given that Construction-Learning Ability was a quasi-ordinal measure, involving ordered levels of difficulty, and was not normally-distributed according to a Shapiro-Wilk test, the Spearman’s method was used. There is a large and significant correlation between Construction-Learning Ability and performance on the TROG (\(\rho = 0.55, p < 0.001\)). Correlations between Construction-Learning Ability and the RAPT and between the TROG and the RAPT were weaker (\(\rho = 0.33, p = 0.244\), and \(\rho = 0.21, p = 0.164\) respectively). To visually assess the relationship between Construction-Learning Ability and the static assessments, scatterplots are shown in Figure 3 along with the line of best fit and 95% confidence intervals.

**Association with combined static assessments of syntactic ability**

Standardised scores from both the TROG and the RAPT were calculated, and the mean obtained, with the results reported in Table 1. The association with Construction-Learning Ability was strong (\(\rho = 0.54, p < 0.001\)), while the association with initial Elicitation Performance was slightly weaker (\(\rho = 0.44, p = 0.002\)).

**Variation in learning rates across constructions**

Further analyses investigated differences in Construction-Learning Ability across the two different constructions. Differences between constructions were investigated via a scatter plot (Figure 4). Each point corresponds to a child, and each axis corresponds to the Construction-Learning Ability for different constructions. While many dots lie close to the line of equivalence, suggesting equal learning rates for both constructions, there are numerous dots below the line suggesting greater difficulties with the *just because*
Construction-learning and language ability

construction. Differences in learning rates for the two constructions were confirmed by a Wilcoxon signed rank test ($V$ (signed rank statistic) = 298, $p = 0.003$).

Further analyses investigated the relationship between Learning Ability and the static assessments on a construction-by-construction basis (3 data points per construction). The *just because* construction was a stronger predictor of performance on the static assessments ($\rho = 0.57$, $p = 0.001$ and $\rho = 0.35$, $p = 0.015$ for TROG and RAPT respectively) than the *what makes* construction ($\rho = 0.44$, $p = 0.002$, and $\rho = 0.18$, $p = 0.211$ for the TROG and RAPT respectively).

*Disentangling construction-learning from STM*

Though the experimental task was designed to investigate construction-learning, the testing paradigm in levels 1 – 3 involved elicited imitation. Therefore the task may conceivably reflect STM abilities rather than construction-learning. To address this possibility, Elicitation Performance at two different time points was investigated: block 1, and block 4 (the final block). This comprises level 4 responses (Elicitation only) which cannot depend on STM as there is no model sentences. It is important to include these responses as they are reflected in the main dependent variable: Learning Ability. It was hypothesised that STM would be more strongly engaged in block 1 than block 4. This is because, if participants are in the process of learning the unfamiliar construction they will be less able to recruit representations in LTM to ‘reconstruct’ the stimulus (Lombardi & Potter, 1992). Consequently, they will be more dependent on verbatim recall skills. By contrast, in the final block, assuming that construction-learning has taken place, repetition of the stimulus will reflect reconstruction of the stimulus from LTM rather than STM capacity. Performance in both blocks 1 and 4 was regressed against the language measures (TROG and RAPT). It was assumed that, according to the Learning Hypothesis, performance in block 4 would demonstrate a stronger association as it is more strongly reflects construction-learning ability.
First block performance was significantly associated with both the TROG (rho = 0.41, p = 0.004) and the RAPT (rho = 0.32, p = 0.028). Though these values are significant, the correlations are weaker than those observed for Construction-Learning Ability (final block performance). A commonality analysis was subsequently employed. This enables one to compare collinear predictors, in order to determine the extent to which they uniquely or jointly accounted for variance in the dependent variable. Such an approach enables one to determine to what extent predictor variables measure the same construct. When the TROG was used as the dependent variable, Construction-Learning Ability (final block performance) was a significant predictor even when first block Elicitation Performance was entered as an additional predictor variable (standardised $\beta = 1.41$, $t = 2.41$, $p = 0.020^*$ for final block performance versus standardised $\beta = 2.72$, $t = 0.82$, $p = 0.416$ for initial block performance).

The results of the commonality analysis are shown in Table 2. Final block performance uniquely predicted $R^2 = 0.15$, compared to $R^2 = 0.01$ for first block performance, with a shared variance of 0.18. The relatively large unique variance indicates that block 4 measures of performance reflect a construct not present in block 1 which is likely to be related to construction-learning ability. When predicting the RAPT, neither variable made a significant contribution when entered simultaneously, indicating that neither made a strong unique contribution (Construction-Learning Ability (i.e. final block performance): standardised $\beta = 0.15$, $t = 0.88$, $p = 0.382$, first block Elicitation Performance; standardised $\beta = 0.24$, $t = 1.24$, $p = 0.222$).

Discussion

The study tested the claim that variation in language ability, ascertained via standardised tests of syntactic ability, reflects differences in the ability to acquire new constructions. This hypothesis is based on arguments, outlined in the introduction, that
learning mechanisms are the locus of individual linguistic variation. To improve ecological-validity, genuine low-frequency constructions were taught. There were substantial differences in construction-learning ability which were associated with static assessments of expressive and receptive syntax (RAPT and TROG), and mean standardised scores. This association was especially strong when the TROG was used as the dependent variable.

The first research goal was to investigate variation in construction-learning abilities. This was, indeed, substantial. A large proportion of children showed hardly any learning of the target constructions, despite receiving input in the maximally supportive condition. Likewise, only children in the top quintile demonstrated strong learning, i.e. progression to the least supportive condition on both structures (see Figure 1). This mirrors the wide variation in language-typical adults when acquiring artificial grammars (Misyak & Christiansen, 2012). Moreover, learning of the what makes you think construction exhibited a bimodal distribution observed in previous AGL studies (Misyak & Christiansen, 2012; Romberg & Saffran, 2013). These findings suggest that a dichotomous view of language, whereby individuals are either typical or impaired, does not do justice to the wide range of abilities observed within the ‘typical’ population.

The second research objective was to investigate the relationship between construction-learning abilities and performance on static language assessments. This relationship was particularly strong when construction-learning ability was used to predict performance on the TROG. The effect size was “large” according to Cohen’s (1988) criteria, and at the high end of the medium range according to Plonsky and Oswald’s (2014) criteria. To further benchmark this correlation we can compare it to correlations between different assessments which ostensibly measure the same construct. The correlation between the TROG, and the concepts and directions subtask of the Clinical Evaluation of Language Fundamentals (Semel, Wiig, & Secord, 2006), both widely-used measures of sentence
comprehension, is 0.56 (Dockrell, Ricketts, Palikara, Charman, & Lindsay, 2010). The correlation between digit and nonword recall, both assessments of short-term memory which cluster together in factor analyses, is 0.38 (Gathercole, Pickering, Ambridge, & Wearing, 2004). Compared against these benchmarks, it would appear that the observed association between construction-learning and the TROG is strong. The unexplained variance could result from fluctuations in attention, or task-specific factors, e.g. articulatory difficulties may impact on elicited imitation performance but not comprehension.

The strong association between construction-learning and the TROG is consistent with the claim, outlined in the literature review, that static syntactic abilities are influenced by construction-learning abilities. It is not certain why construction-learning ability predicted the TROG far better than the RAPT. One possibility is that the weaker correlation with the RAPT is a result of weaker reliability. Though there are no official test-retest reliability measures for the RAPT, because it is an open-ended assessment, dependent on how the children interpret the pictures, one would predict limited reliability. This may also have weakened the association between the RAPT and TROG. Another possibility is that findings are related to the hypothetical ‘primacy’ of comprehension: the idea that linguistic structures and items are acquired in the receptive modality, with production coming later (Clark & Hecht, 1983; though see Pickering & Garrod, 2007 for an alternative view). Interestingly, the stronger association between Elicitation Performance and the TROG cannot be explained in terms of task similarity, as the former involves production mechanisms which are not involved in the latter.

One potential limitation of the learning paradigm is that it depended heavily on the elicited imitation method, which is likely to engage STM. Consequently, performance may reflect STM capacity rather than construction-learning abilities. However, experimental work suggests that elicited imitation depends minimally on STM, and instead taps representations
in LTM (Lombardi & Potter, 1992). Furthermore, the measure of construction-learning ability is derived from the final block, allowing participants to acquire syntactic representations to support recall. Such representations are presumably stored in LTM as they are built up over successive representations accrued over a time period which greatly exceeds STM span. Additionally, good learners moved to the least supportive level (elicitation) which did not tap recall ability. Commonality analyses with the TROG as the dependent variable supported the conjecture that final block performance reflects a learning-related factor, given that it explained substantial unique variance (0.15). In contrast, analyses with the RAPT as the dependent variable did not identify a separate factor in the final block. Nonetheless, it could be argued that the TROG acts as a more valid and reliable dependent variable, given its stronger correlations with the measures of elicitation performance.

A further aim of the study is to investigate the feasibility of using unfamiliar structures to test construction-learning. As argued in the literature review, it is difficult to create novel constructions which are sufficiently language-like. For example, some studies have taught novel constructions with unusual word orders whose meanings are expressible using existing constructions (Wonnacott et al., 2012). These violate the principle of contrast, and consequently children are likely to be aware of their artificiality. Using unfamiliar constructions avoids this difficulty, but introduces other difficulties, e.g. controlling for prior exposure. The data indicate that this approach was relatively successful. Firstly there was substantial variation in learning trajectories, indicating that the task tapped variation in construction-learning abilities. We would not predict such variation if the construction was already known by most of the children. Secondly, the concurrent validity of the construction-learning measure, its significant association with other language assessments, again suggests that it tapped a genuine factor related to individual difference.
Naturally, with this method, it was impossible to control for prior exposure and this limitation should be borne in mind. There is a trade-off between ecological validity and experimental control, with the current study increasing the former at the expense of the latter. One approach to this trade-off is to investigate a phenomenon using a variety of different methodologies, some of which emphasise control while others emphasise ecological validity. Convergent findings from these different approaches would provide stronger support for the hypothesis in question.

Further exploratory analyses investigated learning as a function of structure. These found that the non-adjacent construction (*just because*) was more difficult to learn within the context of the experimental task, and, moreover, was more strongly associated with the static assessments of syntactic ability. These are consistent with studies finding that the learning of non-adjacent artificial grammars, e.g. AXB grammar, is strongly associated with overall syntactic abilities (Hsu et al., 2014; Misyak & Christiansen, 2012). As a caveat, it should be noted that the constructions differed in other ways which were not controlled for. For example, the *what makes* construction contained more invariant material than the *just because* construction given that dependent clauses always started with ‘you think’. Consequently more closely-controlled studies, employing novel stimuli, are needed to more reliably explore which formal properties of constructions affect learning.

The findings of the study have implications for language assessment. Over the past couple of decades, many clinical researchers have espoused ‘dynamic’ assessments, which measure children’s ability to learn over time rather than ‘static’ assessments, which measure children’s performance at a particular moment in time (e.g. Camilleri & Botting, 2013). Almost all standardised assessments fall into the latter category. Dynamic assessments may be better at predicting response to intervention than static assessments (Hasson, Dodd, & Botting, 2012), and may more accurately measure underlying abilities as they are less
culturally-biased (Peña, Iglesias, & Lidz, 2001). The current study supports the validity of Dynamic Assessment, as it suggests that measures of learning are strongly associated with static assessments.

The study was limited in scope, focusing on the association between construction-learning and static performance. Future studies could introduce a test battery to independently measure important cognitive constructs, e.g. STM, chunking, and AGL abilities. Moreover, the observed correlations do not imply causation. For example, construction-learning may be associated with static syntactic abilities not because one drives the other, but because both are dependent on a third factor, e.g. STM. A longitudinal study may help us to determine causality via investigation of cross-lagged correlations. For example, if Construction-Learning Ability at Time 1 predicts grammatical comprehension at Time 2 more strongly than grammatical comprehension at Time 1 predicts Construction-Learning Ability at Time 2, we would have stronger evidence that Construction-Learning Ability affects grammatical comprehension. In addition, future studies should test long-term retention, which is a crucial component of learning and was missing from the current study.

The current study provided support for the Learning Hypothesis, in that an association was observed between construction-learning abilities and performance on static assessments of syntactic abilities, with a particularly strong association observed for the TROG (rho = 0.55, p < 0.001). This hypothesis was formulated to explain individual variation in language abilities without depending on the traditional competence-performance distinction, which has been challenged by researchers from Cognitive Linguistic and usage-based perspectives (Ibbotson, 2013). As argued in the introduction, the competence-performance dichotomy enables us to explain individual differences in terms of either limited competence, or poor performance. By contrast, the usage-based account, which views competence and performance as indivisible, does not allow this option. Moreover, an appeal to input
frequency cannot resolve this dilemma given that some children struggle to acquire language from a rich and varied input (Leonard, 2014). To resolve this issue, it was proposed the variation in language-learning ability could be the key driver of individual differences. In support of this hypothesis, there is substantial converging evidence that children whose language difficulties are diagnosed via static assessments also have difficulties learning new linguistic items (e.g. Hsu et al., 2014).

Though finding of a close link between construction-learning abilities and performance on static syntactic assessments provides initial support for the Learning Hypothesis, it is acknowledged that the study is correlational, and that causation is difficult to establish. As argued above, a longitudinal study may provide much stronger evidence for a causal relationship. In addition, further research is necessary to identify psycholinguistic mechanisms which may underpin both construction-learning and performance on static syntactic assessments.

Acknowledgements

The authors would like to thank the staff and teachers at the schools we visited, and the parents for providing consent. We would also like to thank the Newcastle University Faculty of Humanities and Social Sciences Faculty Research Fund scheme for supporting the study. Finally, thanks to the anonymous reviewers for their insightful input.


Dabrowska, E., & Street, J. (2006). Individual differences in language attainment:
Comprehension of passive sentences by native and non-native English speakers.
*Language Sciences, 28*(6), 604–615.


https://doi.org/10.1044/2015_JSLHR-L-14-0042


Construction-learning and language ability