

Heike Behrens, Karin Madlener and Katrin Skoruppa¹

The role of scaffolding in children's questions: Implications for (preschool) language assessment from a usage-based perspective

DOI 10.1515/gcla-2016-0015

Abstract: This article outlines a range of theoretical, empirical, and practical desiderata for the design of (preschool) language assessments that follow from recent insights into language development from a cognitive-linguistic and usage-based perspective. To assess children's productive communicative abilities rather than their ability to judge the acceptability of complex sentences in isolation is a new perspective in language testing that requires theoretical motivation as well as operationalizable criteria for judging the appropriateness of children's language productions, and for characterizing the properties of their language command. After a brief review of the basic rationale of current strands of preschool assessment in Germany (Section 2), the fundamental usage-based assumptions regarding children's developing linguistic competence and their implications for the design of preschool language diagnostics are characterized (Section 3). In order to assess children's language production, in particular its flexibility and productivity in context, a test environment needs to be created in which children are allowed to use a certain range of language in meaningful contexts. Section 4 thus zooms in on the central question of scaffolding. Section 5 presents corresponding corpus evidence for adult strategies of prompting children to elaborate their answers and for typical child responses. Sections 6 and 7 discuss the corpus-based findings with respect to their implications for the design of (preschool) language assessment and point to further challenges.

Keywords: usage-based approach, language assessment, scaffolding, prompting, productivity, pragmatic adequacy

¹ Authors in alphabetical order. All authors contributed equally to this paper.

Heike Behrens: University of Basel. E-mail: heike.behrens@unibas.ch

Karin Madlener: University of Basel. E-mail: karin.madlener@unibas.ch

Katrin Skoruppa: Université de Neuchâtel. E-mail: katrin.skoruppa@unine.ch

1 Introduction: The importance of (oral) language proficiency for academic success

In Germany, students with a heritage language background (who currently account for more than 20% of the students [cf. Reich 2005: 126]) generally display less academic success than those whose first language is German. These differences are more important in Germany than in most of the other OECD countries (Stanat et al. 2010: 200) and they largely hold after controlling for socioeconomic factors (Stanat et al. 2010: 220–222). Bilingual students are significantly underrepresented in High Schools in Germany and significantly overrepresented in basic secondary education and among the population without any certificate of secondary education (Jeuk 2010: 17–20 [cf. Lakshmanan 2014: 1494 for more general concerns regarding the “over-referral of linguistically and culturally diverse children among those diagnosed for special education classes”]).

Pfaff (2009: 213) argues that the main reason for multilingual children’s educational disadvantage is the fact that “German proficiency is essential to facilitate transitions from preschool to school”. In other words, language competence is a main prerequisite for academic success. Bilingual children are typically found to be inconspicuous with regard to their oral language, that is, their basic interpersonal communication skills, but they display problems with written language and literate registers, that is, the so-called cognitive academic language proficiency (cf. Cummins 2008). Oral language competence can also be seen as a precursor skill for reading and writing. If learning to read and write requires a reconfiguration of oral language skills (Portmann-Tselikas 2011), preschool language development predicts reading and writing, and reading and writing skills in turn predict academic success later on (cf. Griffin et al. 2004; Hart and Risley 1995; Labrell et al. 2013; Lewis et al. 2000).

2 Preschool language assessment

Preschool language assessment aims at the early diagnostics of developmental risks and can lead to additional language training for children who significantly lag behind the expected competence level at age 4;6 or 5;0 (i.e., a year before school starts in Germany). It targets both (simultaneously and successively) bi-/multilingual children with German as a second language and monolingual children with German as their first language who run the risk of language-based educational disadvantages.

2.1 Diagnostic methods and aims

A diagnostic tool may either answer the question ‘who needs help?’ and identify children who will probably not be able to participate in learning activities in school. This type of assessment is commonly called *selektionsdiagnostisch* ‘selection diagnostics’. Or it may aim at the identification of the specific problems in these children’s language development, which should then be subject to additional interventions. This type of assessment is called *förderdiagnostisch* ‘special needs diagnostics’ (cf. Kleissendorf and Schulz 2010; Settineri 2012a). Diagnostic methods include questionnaires, analyses of spontaneous speech (e.g., so-called *profile analyses*), and tests. Tests are assumed to be standardized and normed, therefore yielding comparable and objective, reliable, and valid data (cf. Settineri 2012b; Reich 2013).

2.2 Problematic theoretical and empirical foundations of current language diagnostics

In Germany, *pediatric* diagnostics at ages 3, 4, and 5 (cf. Voet Cornelli et al. 2011) rely heavily on parental reports, which makes the assessment notoriously problematic for L2 children. For *pedagogical* preschool diagnostics in kindergartens, an array of tests and diagnostics has been developed in the past 15 years. Yet most of these assessments are not communicatively embedded but take the form of discrete-item tests that elicit language in highly formal settings with the aim of testing the constructional repertoire, not its contextualized use. Such discrete-item approaches rely on models of acquisition that assume that “L2 proficiency transpires by internalizing simple, discrete units of the L2 before acquiring complex sequences, the accumulation of which constitute[s] proficiency” (Purpura 2016: 194).

In contrast, for instance, to a fair amount of current foreign-language assessments targeting older learners’ linguistic abilities (cf. Purpura 2016: 193–200), German preschool assessments tend to be strictly trait-based and context-free, lacking basic characteristics such as skill-integration (i.e., they target different skills in separate sections; see DOBINE by Quasthoff, Fried, Katz-Bernstein, Lengning, Schröder and Stude 2011 for an exception) and the meaningful grounding in daily-life tasks or in purpose-driven scenarios drawing on critical features of the setting, including cognitive complexity, communicative goals, and socio-interactive aspects of language such as discursive practices, the co-construction of meaning, and the pragmatics of interpersonal relationships between the interlocutors (cf. Purpura 2016: 193–200; see also Lakshmanan

2014). Critically, they only account for children’s very first reactions to stimuli and do not allow for elaboration triggered by follow-up clarification requests or prompts (i.e., precisely the co-construction of discourse). In general, they rest on a deficit view, as they do not distinguish available knowledge resources from strategic resources for communication and learning, for example, pragmatic abilities and language management (cf. Bredel 2005; Quasthoff et al. 2011). Finally, most currently available tests evaluate students’ performance according to monolingual written language standards (cf. Bredel 2005), which “may be inappropriate when assessing bilingual children” (Lakshmanan 2014: 1495); as such, they do generally not allow for different degrees of acceptability and different ways of problem solving.

A final concern is the data that are elicited and used for assessment. Target domains tend to be tested on a small range of exemplars in isolated sentences, and scoring is typically based on syntactical or lexical correctness. Usage-based analyses, in contrast, focus on learners’ degree of constructional productivity, for example, their ability to freely use particular constructions with a range of lexical items across a range of contexts, and on utterance complexity in terms of information density, the relation of the number of propositions to the number and nature of clause elements.

3 Implications for language assessment from a usage-based perspective

How can current theories and empirical findings in the domain of language acquisition inform and help to enhance the design of (preschool) language assessment tools? In the following, we take a cognitive linguistic, usage-based approach that builds on the insights from intense language learning studies over the past 20 years and, more specifically, on what corpus analyses can reveal about naturalistic language development in the critical age range (3;0 to 5;0).

3.1 Language development from a usage-based perspective

Usage-based approaches posit that “[l]anguage knowledge is constituted by a structured inventory of constructions” (Ellis 2015: 50; cf. Goldberg 2003, 2009). Conventionalized, routinized form-meaning pairings at different levels of complexity and abstraction are used for specific communicative purposes, such as words, multiword units, and more or less generalized schemas. Usage-based approaches additionally posit that “knowledge of language is ‘constructed’ on

the basis of the input together with general cognitive, pragmatic and processing constraints” (Goldberg 2009: 93–94) during meaningful, socially contextualized interactions (cf. Behrens 2009; Quasthoff 2015; Tomasello 2003). Language acquisition thus consists of gradually extending and consolidating the constructional repertoire as a rich, densely interconnected network, based on recurrent form-meaning mappings segmented from the input. In addition, speakers need to automatize access to constructions as mental routines during online processing for comprehension and production.

Importantly, usage-based approaches assume that “[l]anguage structure emerges ontogenetically from usage in particular contexts. Development is slow and gradual, moving from an initial reliance on concrete items to more abstract linguistic schemata” (Ellis 2015: 50, 63; cf. Ellis and Wulff 2015a, 2015b; Ortega 2015; Tomasello 2003). Over time, item-general constructions may be abstracted, as need arises (cf. Wray 2002), due to the recognition of both the recurring similarities in (more lexically specific) frequent form-meaning mappings and aspects of systematic variation in forms and functions across utterances and situations.

Two aspects of usage-based acquisition research are relevant in this context: First, language development is grounded in meaningful communication. Particularly between ages three and five, increasingly complex and increasingly explicit child productions can be observed at the levels of lexical, utterance, and discourse-level constructions; for example, children start to ask more complex questions, answer more complex or abstract questions, and provide more complex and elaborate answers to questions (cf. Labrell et al. 2013). Pragmatic skills are reflected in the children's ability to use elliptical structures in order to distinguish between given and new referents, and to rephrase and paraphrase their utterances if need arises.

Second, children's developing language is partially formulaic. Children may thus use complex, but unanalysed linguistic units (formulaic utterances, collocations, chunks, etc.). This means that learner language can be correct at the surface level, although learners have not yet analysed the internal structure of the utterance, or the full paradigmatic variability that these constructions allow (cf. Wray 2002). The production of a particular lexeme, morpheme, or syntactic construction does thus not provide evidence that children really master the underlying semantic or grammatical properties. Thus, a lot of attention is devoted to test the degree of productivity, either in experimental settings or by analysing the variability across several instances of naturalistic speech production (Behrens 2016). Children's increasing lexical inventories (of word-like and multiword units) and their growing degrees of freedom in using more abstract constructions (with a broader range of lexical items, with new and novel lexical items) are indicators of their actual language skills.

These findings may have strong implications for language testing: If children's linguistic productivity in meaningful and motivated conversational settings is to be assessed, they need to be given several chances to produce what they can. This proposal is in line with results from studies on interaction with children.

3.2 Language assessment from a usage-based perspective

Language assessment consists in gathering observable evidence in specific test situations, but results in inferences about individuals' capabilities to use language beyond the test situation (cf. Callies et al. 2014: 71; Mislavy and Yin 2009: 253; Purpura 2016: 191–192). The test situation should thus create a rich external context that elicits specific language usage events, where these usage events are interpretable, as a function of the task complexity, as predictors of both the testees' current overall language competence, namely in real-life situations, and their further potential for development or else, their developmental risks (Mislavy and Yin 2009: 253).

Usage-based approaches do not look at language in terms of isolated abilities, but at contextualized language use across different levels of form-meaning mappings (phonological, lexical, abstract/grammatical, discourse). Usage-based language assessment aims to identify children's strategies in dealing with complex communicative situations, by investigating not only their possible constructional inventory, but also their ability to put these constructions to use in specific communicative settings. Hence the primary focus will be on pragmatic adequacy and informativeness, not on eliciting a particular syntactic construction as the only "correct" response. As such, usage-based approaches to language assessment are skill-integrated, task-centered, scenario-based, and sociointeractional (as opposed to trait-based and context-free, cf. Purpura 2016: 193–200). Language assessment should thus consist of contextualized and meaningful tasks, that is, engage children in meaningful interaction rather than testing word or grammar knowledge in a discrete-item fashion. It should target the (depth, breadth, and productivity of) children's constructional inventory as it is put to use in relevant task contexts for communicative purposes (cf. also Reich 2005: 421).

In order to assess a child's range and extent of knowledge (and his/her ability to access this knowledge for online language processing), a usage-based approach will analyse the degree of repetition and variation across target constructions and task contexts. Since language development is slow and piece-

meal, single occurrences of specific structures do not constitute evidence for a child's command of a corresponding more abstract schema (or category). Thus, a substantial range of systematically varied target contexts for a reliable estimate of children's constructional productivity and variability are needed: Is the child able to vary his/her verbal expressions or does s/he stick to the same verbal routine across a selection of contexts? Is s/he willing and able to increasingly add detail when the situation requires him/her to specify new contrasts, or does the level of differentiation and specificity remain the same?

In order to estimate children's language skills along these dimensions, they will not be compared to an abstract adult written norm, but to a "representative" range of naturalistic and/or elicited child performance data. As a baseline of comparison, we need to establish attested and pragmatically adequate ways to deal with the communicative situations/stimuli selected for assessment: Are there particularly frequent or basic structures that are (or should be) mastered by all children at this age? Which are the verbal means that are only mastered by the most advanced children? What is the range of constructions used, and how do they differ in their degree of adequacy and informativeness? Which strategies do children use who do not use the "typical" constructions? And are children who do not yet master the typical constructions just delayed (and thus typical for an earlier stage of typically developing L1 children), or do they use different constructions?

Finally, as usage-based approaches assume that "[l]anguage learning takes place within the framework of social interaction" (Nelson 1985: 109) and that "[l]anguage in its conversational settings does social work" (Clark 2009: 50), language assessment will not (exclusively) elicit monologic, but rather dialogic child data, that is, child responses in a largely naturalistic (preschool) setting of language use. As for such naturalistic settings, children have expectations regarding the discursive collaboration of their (adult) interlocutors; for example, multilingual children have been shown to produce shorter utterances and more fragments when talking to adults (as compared to child-child discourse, cf. Röhner and Oliva Hausmann 2011) precisely because they assume that adult experts will be collaborative and scaffold their discourse (see below). As a function of the difficulty of individual tasks, the assessment will thus provide multiple sources for the co-construction of discourse through prompting and scaffolding, namely through clarification requests or other types of follow-up questions, which are common in naturalistic adult-child interactions and should allow the children to perform at their best. It will possibly also take into account children's language management skills and their metalinguistic sensitivity, for example, their ability for self-repair and self-evaluation.

4 The importance of scaffolding

The general goal is to assess language in meaningful communication and thus to avoid some of the problems of discrete-item testing, where scoring is typically based on a predefined complete answer to a specific question. If usage-based insights as outlined above are to be applied, we are in the realm of conversation, where discourse is co-constructed and evolves over sequences of turns (e.g. Quasthoff 2015). For example, the listener can signal whether the information provided is sufficient, or whether more information is wished for. In child language development there is a long tradition of looking at such interactional cues in the tradition of research on child-directed speech (Snow 1977). Adults provide subtle cues when children's productions are not correct (Chouinard and Clark 2003) and they help children to learn language by expanding or elaborating on their utterances (so-called scaffolding).

Scaffolding is “a social mediation involving two people, and is performed by a person who is an expert” (Cook 2008: 229). As such, it “enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts” (Wood et al. 1976: 90, qtd. in Clark and Graves 2005: 571; cf. also Clark 2009: 295). As Graves and Graves point out, scaffolding may not only help children to complete tasks that they would be unable to complete alone, but also help them to “better complete a task, to complete a task with less stress or in less time” (2003: 30).

The adult expert is thus one step ahead of the child (in terms of Vygotsky's (1978) *zone of proximal development*) and systematically uses his or her greater skills or knowledge to construct an interactional level that allows the child to participate as an interlocutor in a challenging context (cf. Hausendorf and Quasthoff 1996: 295). The expert interlocutor counterbalances the skill difference in order to facilitate a joint communicative success (Quasthoff 2015: 304). S/he typically focuses the child's attention onto the most relevant aspects of the given situation or task, reduces the information density of the task or the child's degrees of freedom by structuring or simplifying the task or the range of options, and consistently points out gaps between current and desired end states in the current problem (solving) space (Mackowiak et al. 2008: 74; cf. Wood et al. 1976). The adults thus adapt to the children's lower skill level and systematically refocus and complement their utterances.

4.1 Scaffolding for learning

Typically, scaffolding is taken to be crucial for learning to take place (cf. Bruner 1978). For instance, Clark and de Marneffe argue that “[a]dults frequently offer

reformulations or construals of child verbs in context, and thereby offer children information about distinct sets of constructions in which a verb form can appear” (2012: 115). They show how children can actually learn specific morphosyntactic contrasts from adults' construals and reformulations of fragmentary child utterances.

4.2 Scaffolding for interaction

However, also moment-to-moment interactional scaffolding for current communicative purposes can be observed, more or less independently of an associated longer-term learning outcome (cf. Quasthoff 2015: 288). So-called *vertical scaffolding* “involves the adult extending the child's language by asking further questions. So in response to the child's utterance ‘cow’, she might [...] ask for an elaboration ‘And what did we see when we went to the farm today?’” (Foley 1994: 101; cf. also Clark and Graves 2005: 572). In this fashion, adult interlocutors can help children to produce “narratives” long before they master the relevant complex constructions in the target language. The contrast between the two following exchanges – the unsuccessful one in (1) with an unfamiliar adult and the successful one in (2) with the mother (reported in Clark 2015: 334) – shows how, based on shared knowledge, the familiar adult in (2) narrows down the relevant information with respect to the pertinent event and prompts the child to “supply just the piece of information needed at that moment” (Clark 2009: 285):

- (1) Child (1;6): Band-aid.
 Observer: Where's your band-aid?
 Child: Band-aid.
 Observer: Do you have a band-aid?
 Child: Band-aid.
- (2) Child (1;6): Band-aid.
 Mother: Who gave you the band-aid?
 Child: Nurse.
 Mother: Where did she put it?
 Child: Arm.

Adult contributions, such as prompts or questions, may thus provide “a partial scaffold for the child's fragmentary contributions” (Clark 2009: 295). As such, they may “often offer a framework into which the child's contributions can be slotted” (Clark 2009: 295). Wessel (2015: 332) points to the potential importance

of the adult's uptake of parts of the child utterance as an anchor for further adjustment and prompting for elaboration, including prompts that help the child to overcome linguistic hurdles as shown in (2) or, for example, to replace vague deictic expressions with more precise non-deictic spatial language. Generally speaking, scaffolding through prompts and clarification requests represents a kind of discursive co-construction. As Clark observed,

[t]he contributions of each participant in conversation rarely have clear boundaries. Speaker and addressee often collaborate to arrive at an expression of the intended meaning. Child and adult may jointly construct a proposition, as when an adult offers a scaffold for the child's contributions in talking about a specific event known to both of them. (Clark 2009: 292)

Scaffolding keeps conversations running and helps to avoid or to repair disruptions (Clark 2009: 304). Scaffolds can take the form of unspecific prompts, also called clarification markers, such as *huh?*, *hum?*, or *what?*, or be repetitions of children's utterances with a raising question intonation (Clark 2009: 304). Furthermore, questions can be used as clarification requests to prompt and facilitate children's verbal elaborations.

As children get older, adult interlocutors ask for increasingly explicit expressions of the meanings intended by children (Clark 2009: 10). Even young children do respond to such clarification requests by repeating themselves or rewording their utterances in order to make themselves understood (cf. Clark 2009: 303–304). Chouinard and Clark (2003) have shown that such repair “often zeroes in on the ‘trouble’ that led the adult to ask for clarification” (Clark 2009: 304).

Based on the rich evidence that scaffolding helps the child to produce language at a higher level of complexity and informativeness than attested in unscaffolded productions, it can be argued that scaffolding and clarification requests should be used when the aim of the test is to assess the range of children's competence, as they may not always produce their ‘best’ utterance right away, and as information is typically distributed over several turns in a naturalistic conversation. We thus propose that scaffolding through prompts and clarification requests may help children to be informative and to identify the relevant piece of information from a range of options in an assessment situation.

In the empirical part of this paper the following questions are investigated:

- a. How do children spontaneously respond to (different types of) adult questions in natural discourse? To what extent do they spontaneously provide syntactically and pragmatically complex and elaborate answers? To this end, question-answer sequences in longitudinal corpora of German are analyzed.

- b. How do adults scaffold child discourse through prompts and follow-up questions? And which prompts and scaffolds will work best? To this end, the conversational sequences that follow the adult prompts are analyzed.

5 Corpus study

If the goal in language assessment is to elicit children's linguistic competence as naturally and comprehensively as possible, the above-mentioned discourse factors need to be taken into account. As stated above, the child should be given the opportunity to show his or her range of communicative competence in a given setting. In the domains of questions and answers, for example, it may be assumed that the child may not give the full answer at once but that the elaboration on the topic may take several turns, in which the interlocutor provides adequate scaffolds that help the child to elaborate. In order to investigate which types of scaffolds create optimal communicative conditions for the child and would thus work best in an assessment environment, we established a corresponding corpus baseline.

5.1 Corpora

We studied one spontaneous speech recording (30–60 minutes) per month for four children aged 4;00 to 4;11 years (two boys, two girls) from German longitudinal corpora in CHILDES (MacWhinney 2000). In doing so, we analysed data from three children (and their caretakers) in the *Rigol* corpus: Sebastian, Pauline, and Cosima. These children are from varied socio-economic backgrounds, and the data include conversations with various partners, such as their mothers, siblings, friends, and the observer, a friend of the family. The fourth data set consists of data from the *Leo* corpus, a high density developmental corpus collected by the MPI for Evolutionary Anthropology (for details, see Behrens 2006). This corpus consists of naturalistic conversations between Leo and his mother, the primary caregiver, who has an academic background.

5.2 Coding

All questions and answers were identified and annotated syntactically and pragmatically, using coding schemes that were inspired by previous work on the

syntax and pragmatics of questions and answers in children (Salomo et al. 2013; van Hekken and Roelofson, 1982; Bishop et al. 2000) and adults (Stivers 2010). In particular, the coders read through the transcripts and identified whether an utterance was part of a question or an answer. For those that were, the coders identified the speaker and the addressee of the utterance (range: target child, sibling, or other adults), the utterance's status in follow-up question chains (if applicable; see Table 3) as well as a range of syntactic and pragmatic categories (see Tables 1 and 2 for answer categories).

Five trained coders annotated the data. Around half of the coding was done by the third author, who also resolved any ambiguous cases highlighted by the other coders in discussion with them. Around 10% of the data were recoded blindly by a different coder, yielding a corrected κ value (Brennan and Prediger 1981) of 0.88 for interrater agreement concerning the conversational partners (speaker and addressee), of 0.94 for interrater agreement for syntactic and of 0.72 for pragmatic categories and follow-up status. These agreement values range from 'substantial' to 'almost perfect', according to Landis and Koch's (1977) classification. Although all questions and answers were coded, the focus will be on questions by the main adults (mothers and observer) directed at the four target children ($n = 5095$) as well as on the children's answers ($n = 3840$) for the purpose of the present analyses.

Table 1: Syntactic categories for (child) answers

| Category | Examples | Translation |
|------------------------|--|---|
| nonverbal interjection | laughing, vocalizing, gestures Hm. Oh. | |
| yes/no | Ja. Hm, nein. | Yes. Hum, no. |
| word | Da. Zitteraal. | There. Electric eel. |
| phrase | Am Baum. Da rein. | At the tree. In(to) there. |
| clause | Liegt auf'm Boden. Das weiss ich nicht. Weil mir kalt ist. | (It's) Lying on the floor. I don't know that. Because I'm cold. |
| complex | Ich weiss, wo das ist. Das ist eine Kuh, die frisst Gras. | I know where it is. It's a cow, it eats grass. |

Table 2: Pragmatic categories for (child) answers

| Category | Explanation | Examples | Translation |
|--------------------|---|--|--|
| exemplary question | | <i>Was ist das?</i> | What is this? |
| off topic | no relation to the question | Ich geh weg. | I'm going away. |
| misunderstanding | answer does not fit the question type | Nein. | No. |
| minimal answer | question acknowledged, but required information not given | Weiss nicht. | Don't know. |
| adequate answer | (only) required information is given | Ein Zitteraal. | An electric eel. |
| elaborate answer | extra information beyond required | Ein Zitteraal. Der ist sehr gefährlich. | An electric eel. It's very dangerous. |

Table 3: Types of follow-up questions

| Category | Explanation | Examples | Translation |
|--------------------|---|-------------------------------|-------------------------|
| exemplary question | | <i>Was ist das?</i> | What is this? |
| unspecific prompt | interjection/one word question to solicit answer/continuation | Hm? Und? | Hm? And? |
| repetition | (near)-identical repetition of the original question | Was ist das denn? | So what is this? |
| reformulation | re-wording of the original question | Das hier, das ist was? | This one, this is what? |
| refocusing | narrowing of the original question (e.g. by specifying one alternative) | Ist das vielleicht ein Pferd? | Is this a horse, maybe? |

5.3 Child answer behavior

The corpus data show that, on the syntactic level, full clauses constitute less than half of the children's answers (Figure 1). Syntactically complex answers that consist of more than one clause are very rare (median 3%). While simple yes/no answers and elliptical answers (i.e., phrases, fragments, or single words) account for roughly a third each. It is noteworthy that elliptical answers are generally communicatively adequate, that is, pragmatically licensed, because they highlight the new information requested (e.g., Observer: *Was ist denn das?* 'What is this?' – Cosima (4;0): *Ein Baum.* 'A tree.').

On the pragmatic level, inadequate answers, such as minimal and off-topic answers, make up more than a quarter of the sample (see Figure 2). Only about

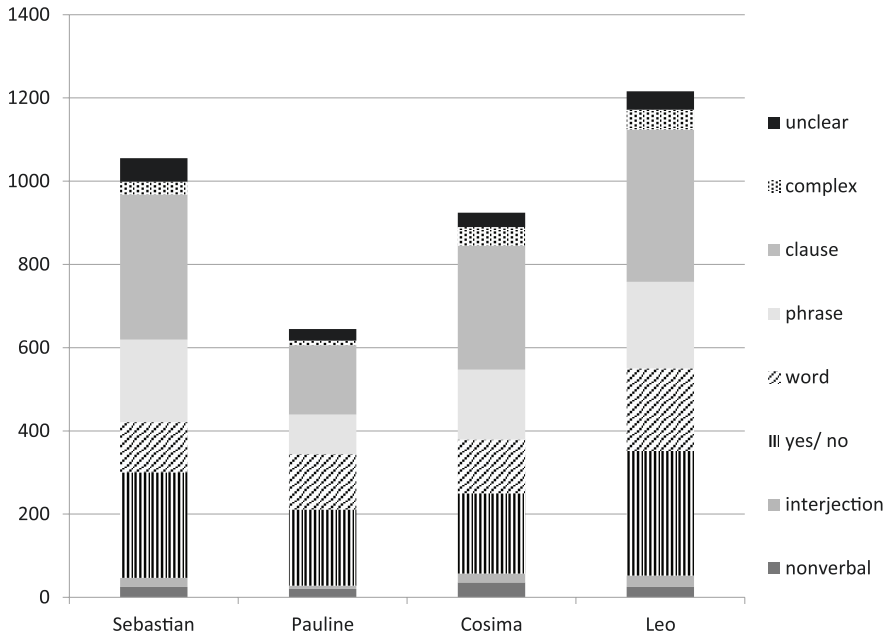


Fig. 1: Answer tokens per syntactic answer category per child

half of the child answers are actually pragmatically adequate, providing the required information. These results are compatible with prior research on preschool children's answer behavior in child-child dyads (according to van Hekken and Roelofson [1982], only 56% of answers amongst five-year-olds are informative and adequate). Thus, naturalistic language tests for this age group should include significantly more questions than would be necessary given a perfect answer behavior.

Elaborate answers to adult questions, that is, answers that spontaneously provide additional information beyond the minimally required information, are rare (median: 10%). Such rich answers often occur when the topic is already well established (i.e., on average after four to seven turns on topic). This highlights the importance of providing good lead-ins for test questions.

Misunderstandings are very rare (87 in total), suggesting that children correctly understand most question categories by preschool age. However, they disproportionately affect (equally rare) alternative questions (e.g., Mother: *Mit Kribbel oder besser mit normalem Leitungswasser?* 'With gas or better with normal tap water?' – Leo [4;3]: *Ja*. 'Yes.'). 12 misunderstandings for 69 alternative questions overall), suggesting that this question category should be avoided in language diagnostics.

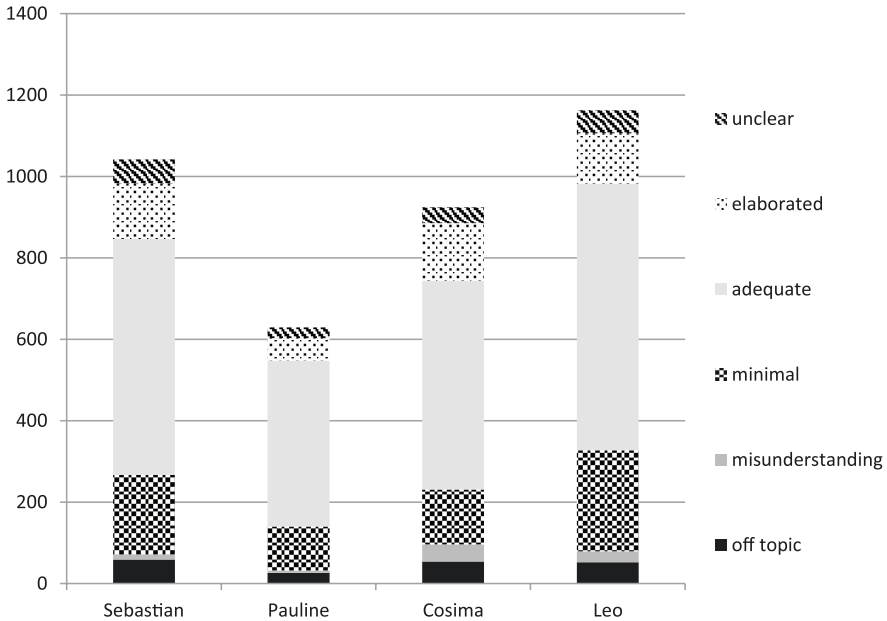


Fig. 2: Answer tokens per pragmatic answer category per child

In addition, a closer look at the Leo data shows that 21% of his mother's questions remain unanswered at first. A large part of these questions can be interpreted as not really requiring an answer (e.g., confirmation questions such as *Da staunst du, hae?* 'Now you are surprised, aren't you?', 32% unanswered). By contrast, questions that truly ask for new information (e.g., *Was willst n gleich bei der Vivien machen?* 'What do you want to do at Vivien's later?') and clarification requests (e.g., *Bitte?* 'Pardon?') rarely remain without an answer (17% and 9% unanswered respectively). Importantly, for 32% of the unanswered questions, Leo's mother produced a follow-up question, which suggests that she does expect an answer in these contexts.

5.4 Scaffolding through follow-up questions

Across the whole corpus, 11.6% of the questions are follow-up questions (591 out of 5095). Figure 3 shows the number of tokens per follow-up question category directed at each child. For each child, refocusing (cf. [3]) accounts for almost half of the follow-up questions, and reformulations are clearly more frequent than repetitions and general prompts.

- (3) Observer: *Was hast du denn gegessen heute?* ‘What did you eat today?’
 Cosima (4;6): 0.
 Observer: *Nudeln mit Tomatensosse?* ‘Pasta with tomato sauce?’
 Cosima: *Nein, Eiernudeln.* ‘No, egg pasta.’

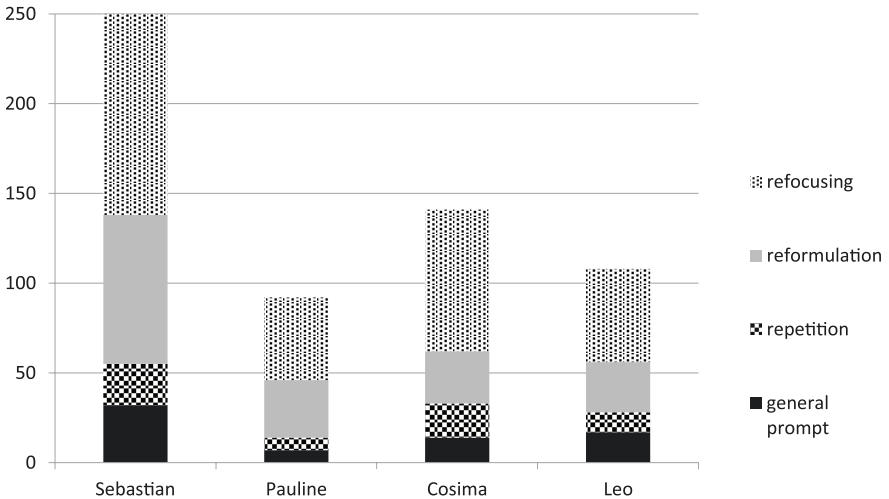


Fig. 3: Tokens of follow-up question categories directed at each child

Two general patterns emerge from these findings: Firstly, all adult interlocutors follow up on questions that are important to them if the child does not answer spontaneously or does not provide an adequate answer to the adult’s question. Children may thus confidently expect adult interlocutors to signal lacking information or otherwise problematic answers and to get a chance to elaborate them in a following turn. They are thus probably not used to verbalize very elaborate, highly informative, syntactically complex answers to adult questions straight-away, as is expected in many language diagnostics.

Secondly, all adults show similar preferences in terms of how to prompt the child to answer questions that remained unanswered at first, to correct misunderstandings, or to elaborate their insufficiently informative or pragmatically inadequate answers. The substantial proportion of refocusing questions shows that adults use clarification requests to clearly point children to the lacking piece of information relevant to the current communicative context and that they scaffold the co-construction of discourse, namely by narrowing down the possible answer options, in order to help children cope with the communicative task.

Children do respond to the adults' prompting and scaffolding intents. All follow-up question chains were perused and it was gauged how well the child had answered the *original* question before the next topic change (see Figure 4). We find that, by the end of follow-up question chains, around 50% of the original questions (to which the children had not provided an answer in the first place) received an adequate, sufficiently informative, or even elaborate response.

Interestingly, the chance to obtain such a satisfactory answer is somewhat higher for reformulations and refocusing questions (ca. 60%), which are also preferred by adult interlocutors (see Figure 3), than for repetitions and un-specific prompts (40% and 50% respectively). A chi square test shows that the distribution of follow-up questions and answer types is different from chance ($X^2 = 23.5$, $p < .001$). In sum, follow-up questions are a promising means to elicit (more elaborate) child answers, in particular those clarification requests that reformulate or refocus (i.e., narrow down) the original question and thus help children to zoom in on the targeted information. Note, however, that a number of refocusing questions in spontaneous speech already contain very strong hints to the desired answer (see example [3] where 'pasta' is already mentioned), which one might want to avoid in language assessment, unless one explicitly investigates Vygotsky's *zone of proximal development*, which defines the next likely step in development as what the child can achieve with support, but not yet without.

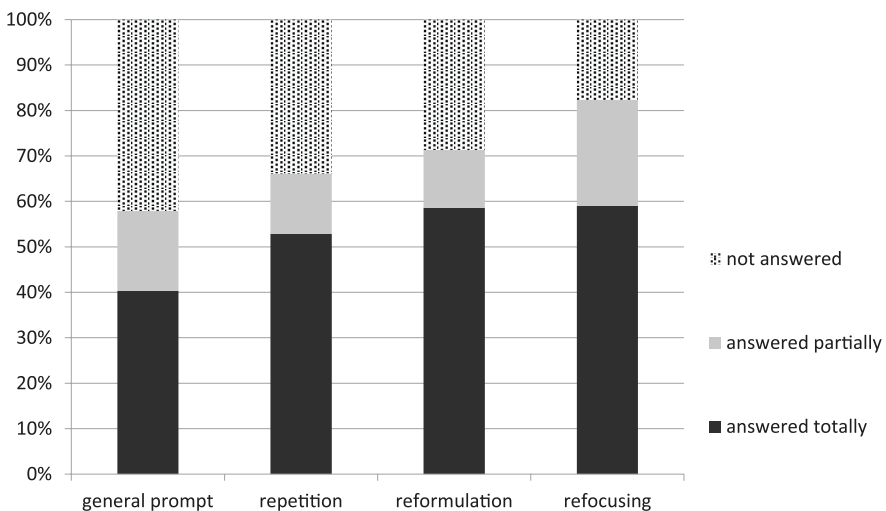


Fig. 4: Answer success with respect to the original question, per first follow-up question category

6 Conclusion: How can (follow-up) questions be used in language diagnostics?

With respect to the nature and types of questions to be used, our data suggest that preschool children understand all major categories, with the exception of alternative questions, which should thus be avoided. Elliptic answers should be expected and receive full scores if they are pragmatically licensed. However, the pragmatic adequacy of children's answers should not be overestimated, since the corpus data show that only about half of them include the requested information in naturalistic adult-child interactions, a figure that is very similar to that found for answers in child-to-child interaction in pre-schoolers (van Hekken and Roe-loffson 1982). Thus, a surplus of questions should be included in language tests in order to make up for the expected proportion of uninformative answers. Furthermore, longer and elaborate answers usually concern well-known and well-established topics, underlining the importance of a good and coherent script leading in to the critical questions.

Finally, the data highlight the importance of following up on insufficient answers, a strategy that can lead to a more adequate response in around 50% of the cases. Interestingly, more targeted reformulation and refocusing has a slightly higher success than general prompts and simple repetitions. Most language tests currently do not allow for any re-iteration on a given test item. Narrative assessments like DOBINE (Quasthoff et al. 2011) and MAIN (Gagarina et al. 2012) form an exception, but they only allow for general, unspecific follow-up prompts (e.g., *hm?* 'hm'; *und?* 'and?') that do not interfere with the child's structuring of the story. If the assessment takes place in the form of targeted elicitation (Eisenbeiss 2009), structured guidance in the form described above may be more efficient. Depending on the linguistic skill tested, it may be possible to provide more specific follow-up questions helping the child to focus on the precise information that is requested, for instance by rephrasing the question or by providing a possible alternative answer. Ultimately, test instructions should not consist of one single question per linguistic task, but of an ordered series of carefully scripted follow-up questions, starting with unspecific prompts and moving on to more specific support for cases of continuous insufficient answers.

7 Outlook

This article constitutes a first step towards an empirically-validated, usage-based assessment of child language production, which aims at eliciting a valid, rep-

representative, and conclusive sample of the child's language when performing at his/her best. A first set of challenges for assessing language in naturalistic interactions was identified and addressed. As outlined above, such an assessment should

- a. be embedded in a communicatively relevant situation, inviting the child to truly interact and share his/her knowledge;
- b. provide more questions than strictly necessary in order to make up for potentially uninformative answers;
- c. allow for scaffolding through follow-up questions in order to help the child to produce more language, and score more than only the first child answer; and
- d. fully accept pragmatically adequate elliptical answers in the scoring system, as elliptic structures actually point to substantial discursive skills regarding the differentiation between given and new (focal) information.

There are (at least) two further challenges for usage-based language tests: First, instead of testing discrete items with a pre-specified 'correct' response, they allow for 'open' responses in order to give children the chance to show the range of their linguistic abilities in the domain under investigation. The resulting richer dataset should also facilitate an assessment of productivity and variability. The assessment situation should thus provide multiple occasions for producing similar constructions in similar but systematically varied situations, and stimulate the child to become increasingly differentiated in his/her utterances, for example, by gradually introducing additional semantic contrasts.

Second, new forms of evaluating multiple open responses have to be devised. The measure of comparison which is needed for the validation of the assessment could be built from a baseline derived from naturalistic data (ideally, from existing monolingual and multilingual learner corpora) that provides information about the use of constructions and lexical items across contexts. If available, such measures of comparison should be based on a representative sample of spontaneous and elicited productions by mono- and bilingual children, matched for age and contact duration, not on an assumed adult written norm. In this way, a strictly dichotomous correct/incorrect scoring can be replaced by graded ratings of the structural complexity and pragmatic adequacy of a given response. Importantly, a usage-based approach will appreciate variation in children's expressions and conceptualizations, and tolerate multiple possible answers, especially with multilingual participants.

Acknowledgements: Part of this research was funded by the project *Sprachstandsermittlung bei Kindern mit Migrationshintergrund*, Daimler und Benz Stiftung, Ladenburg, Germany. The project was initiated by Wolfgang Klein, and

is headed by Jörg Roche. Further project members include Stefanie Habertzettl, Marcus Hasselhorn, Dirk Ifenthaler, Moiken Jessen, Natalia Kapica, Gabriele Kecker, Giulio Pagonis, Maike Schug, Elisabetta Terrassi-Haufe, Frank Thissen, and Nicole Weidinger. We would like to thank our students for data coding: Cornelia Fanger, Isabel Lüthi, Ariane Schwaiger, and we would like to express our special gratitude to Julia Voegelin. Thanks also to the discussants at the international conference *Thinking, Doing, Learning 2* (Groningen, June 2015) for their critical feedback. The usual disclaimers apply.

References

- Behrens, Heike. 2006. The input-output relationship in first language acquisition. *Language and Cognitive Processes* 21. 2–24.
- Behrens, Heike. 2009. [Usage-based and emergentist approaches to language acquisition. *Linguistics* 47\(2\). 383–411.](#)
- Behrens, Heike. 2016. The acquisition of grammatical categories. In Edith L. Bavin & Letitia R. Naigles (eds.), *The Cambridge handbook of child language*, 2nd edn. Cambridge: Cambridge University Press. 250–270.
- Bishop, Dorothy V. M., Janet Chan, Catherine Adams, Joanne Hartley & Fiona Weir. 2000. Conversational responsiveness in specific language impairment: Evidence of disproportionate pragmatic difficulties in a subset of children. *Development and Psychopathology* 12. 177–199.
- Bredel, Ursula. 2005. Sprachstandsmessung: Eine verlässene Landschaft. In Konrad Ehlich (ed.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bundesministerium für Bildung und Forschung. 77–119.
- Brennan, Robert L. & Dale J. Prediger. 1981. Coefficient κ : Some uses, misuses, and alternatives. *Educational and Psychological Measurement* 41. 687–699.
- Bruner, Jerome S. 1978. The role of dialogue in language acquisition. In Anne Sinclair, Robert J. Jarvella & Willem J. M. Levelt (eds.), *The child's concept of language*. New York: Springer-Verlag. 241–256.
- Callies, Marcus, María Belén Díez-Bedmar & Ekaterina Zaytseva. 2014. Using learner corpora for testing and assessing L2 proficiency. In Pascale Leclercq, Amanda Edmonds & Heather Hilton (eds.), *Measuring L2 proficiency: Perspectives from SLA*. Bristol et al.: Multilingual Matters. 71–90.
- Chouinard, Michelle M. & Eve V. Clark. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30(3). 637–669.
- Clark, Eve V. 2009. *First language acquisition*. 2nd edition. Cambridge et al.: Cambridge University Press.
- Clark, Eve V. 2015. Common ground. In Brian MacWhinney & William O'Grady (eds.), *The handbook of language emergence*. Chichester: Wiley Blackwell. 328–353.
- Clark, Eve V. & Marie-Catherine de Marneffe. 2012. Constructing verb paradigms in French: adult construals and emerging grammatical contrasts. *Morphology* 22. 89–120.

- Clark, Kathleen F. & Micheal F. Graves. 2005. Scaffolding students' comprehension of text. *The Reading Teacher* 58(2). 570–580.
- Cook, Vivian. 2008. *Second language learning and language teaching*, 4th edn. London: Routledge.
- Cummins, Jim. 2008. BICS and CALP: Empirical and theoretical status of the distinction. In Brian V. Street & Nancy Hornberger (eds.), *Encyclopedia of Language and Education*, 2nd edn. New York: Springer Science & Business Media LLC. 71–83.
- Eisenbeiss, Sonja. 2009. Contrast is the name of the game: Contrast-based semi-structured elicitation techniques for studies on children's language acquisition. *Essex Research Reports in Linguistics* 57(7).
- Ellis, Nick C. 2015. Cognitive and social aspects of learning from usage. In Teresa Cadierno & Søren W. Eskildsen (eds.), *Usage-based perspectives on second language learning*. Berlin & Boston: De Gruyter Mouton. 49–73.
- Ellis, Nick C. & Stefanie Wulff. 2015a. Second language acquisition. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*. Berlin & Boston: De Gruyter Mouton. 409–431.
- Ellis, Nick C. & Stefanie Wulff. 2015b. Usage-based approaches to SLA. In Bill VanPatten & Jessica Williams (eds.), *Theories in second language acquisition: An introduction*, 2nd edn. New York & London: Routledge. 75–93.
- Foley, Joseph. 1994. Key concepts in ELT: Scaffolding. *ELT Journal*, 101–102.
- Gagarina, Natalia, Daleen Klop, Ute Bohnacker, Kuola Tantele, Sari Kunnari, Taine Välimaa, Ingrida Balciuniene & Joel Walters. 2012. MAIN: Multilingual assessment instrument for narratives. *ZAS Papers in Linguistics* 56.
- Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7. 219–224.
- Goldberg, Adele E. 2009. The nature of generalization in language. *Cognitive Linguistics* 20(1). 93–127.
- Graves, Michael F. & Bonnie B. Graves. 2003. *Scaffolding reading experiences: Designs for student success*. Norwood, MA: Christopher-Gordon.
- Griffin, Terri M., Lowry Hemphill, Linda Camp & Dennis Palmer Wolf. 2004. Oral discourse in the preschool years and later literacy skills. *First Language* 24(2). 123–147.
- Hart, Betty & Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes Publishers.
- Hausendorf, Heiko & Uta Quasthoff. 1996. *Sprachentwicklung und Interaktion: Eine linguistische Studie zum Erwerb von Diskursfähigkeiten bei Kindern*. Wiesbaden: Westdeutscher Verlag.
- Hekken, Suus M. J. van & Wim Roelofsen. 1982. More questions than answers: A study of question–answer sequences in a naturalistic setting. *Journal of Child Language* 9. 445–460.
- JeuK, Stefan. 2010. *Deutsch als Zweitsprache in der Schule. Grundlagen, Diagnose, Förderung*. Stuttgart: Kohlhammer.
- Kleissendorf, Barbara & Petra Schulz. 2010. Sprachstandserhebung zweisprachiger Kinder in der Praxis am Beispiel Hessen. In Martina Rost-Roth (ed.), *DaZ-Spracherwerb und Sprachförderung Deutsch als Zweitsprache. Beiträge aus dem 5. Workshop Kinder mit Migrationshintergrund*. Freiburg i.Br.: Filibach. 143–160.
- Labrell, Florence, Georges Dellatolas, Sophie Griffini & Clémence C. Lalande. 2013. Short expressive language evaluation by French teachers during children's fourth year (SELEFT4). *Revue européenne de psychologie appliquée* 63. 25–32.

- Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33. 159–174.
- Lakshmanan, Usha. 2014. Bilingual assessment. In Antony J. Kunnan (ed.), *The companion to language assessment*. Malden et al.: John Wiley & Sons, Inc. 1492–1509.
- Lewis, Barbara A., Lisa A. Freebairn & H. Gerry Taylor. 2000. Academic outcomes in children with histories of speech sound disorders. *Journal of Communication Disorders* 33. 11–30.
- Mackowiak, Katja, Gerhard W. Lauth & Ralf Spieß. 2008. *Förderung von Lernprozessen*. Stuttgart: Kohlhammer.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*, 3rd edn., vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mislevy, Robert J. & Chengbin Yin. 2009. If language is a complex adaptive system, what is language assessment? *Language Learning* 59, supplement 1. 249–267.
- Nelson, Katherine. 1985. *Making sense: The acquisition of shared meaning*. New York: Academic Press.
- Ortega, Lourdes. 2015. Usage-based SLA: A research habitus whose time has come. In Teresa Cadierno & Søren W. Eskildsen (eds.), *Usage-based perspectives on second language learning*. Berlin & Boston: De Gruyter Mouton. 351–370.
- Pfaff, Carol. 2009. Parallel assessment of oral and written text production of multilinguals: Methodological and analytical issues. In Bernt Ahrenholz (ed.), *Empirische Befunde zu DaZ-Erwerb und Sprachförderung. Beiträge aus dem 3. Workshop Kinder mit Migrationshintergrund*. Freiburg i.Br.: Filibach. 213–233.
- Portmann-Tselikas, Paul. 2011. Erzählen und Textkompetenz. Linguistische Aspekte des Erzählens. *Informationen zur Deutschdidaktik* 3. 53–62.
- Purpura, James. 2016. Second and foreign language assessment. *Modern Language Journal* 100 (supplement 2016). 190–208.
- Quasthoff, Uta. 2015. Ko-Konstruktion in Erwachsenen-Kind Interaktionen: *membership* und der Erwerb von sprachlichen Kompetenzen. In Ulrich Dausendschön-Gay, Elisabeth Gülich & Ulrich Krafft (eds.), *Ko-Konstruktionen in der Interaktion. Die gemeinsame Arbeit an Äusserungen und anderen sozialen Ereignissen*. Bielefeld: Transcript Verlag. 287–312.
- Quasthoff, Uta, Lilian Fried, Nitza Katz-Bernstein, Anke Lengning, Anja Schröder & Juliane Stude. 2011. (Vor-)Schulkinder erzählen im Gespräch: Kompetenzunterschiede systematisch erkennen und fördern. *Das Dortmunder Beobachtungsinstrument zur Interaktions- und Narrationsentwicklung (DO-BINE) und der Dortmunder Förderansatz (DO-FINE)*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Reich, Hans. 2005. Forschungsstand und Desideratenaufweis zu Migrationslinguistik und Migrationspädagogik für die Zwecke des „Anforderungsrahmens“. In Konrad Ehlich (ed.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bundesministerium für Bildung und Forschung. 120–169.
- Reich, Hans. 2013. Sprachstandserhebungen, ein- und mehrsprachig. In Bernt Ahrenholz & Ingelore Oomen-Welke (eds.), *Deutsch als Zweitsprache*, 3rd edn. Baltmannsweiler: Schneider Verlag Hohengehren. 420–429.
- Röhner, Charlotte & Andrés Oliva Hausmann. 2011. Zweitsprachliche Produktivität von Migrantenkindern im Übergang vom Kindergarten zur Grundschule. In Bernt Ahrenholz (ed.), *Deutsch als Zweitsprache. Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund*. Freiburg i.Br.: Filibach. 75–93.

- Salomo, Dorothé, Elena V. M., Lieven, Elena & Michael Tomasello. 2013. Children's ability to answer different types of questions. *Journal of Child Language* 40. 469–491.
- Settinieri, Julia. 2012a. Zur Messgüte sprachstandsdiagnostischer Verfahren im Vorschulalter: Erste Ergebnisse der Korrelationsstudie Sismik/Seldak – Delfin 4. In Bernt Ahrenholz & Werner Knapp (eds.), *Sprachstand erheben – Spracherwerb erforschen. Beiträge aus dem 6. Workshop Kinder mit Migrationshintergrund*. Stuttgart: Filibach bei Klett. 35–51.
- Settinieri, Julia. 2012b. Möglichkeiten und Grenzen der Prüfung konvergenter Validität sprachstandsdiagnostischer Verfahren. In Bernt Ahrenholz (ed.), *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren*. Berlin et al.: De Gruyter. 325–348.
- Snow, Catherine E. 1977. Mothers' speech research: From input to interaction. In Catherine E. Snow & Charles A. Ferguson (eds.), *Talking to children: Language input and acquisition*. London: Cambridge University Press. 31–49.
- Stanat, Petra, Dominique Rauch & Michael Segeritz. 2010. Schülerinnen und Schüler mit Migrationshintergrund. In Eckhard Klieme, Cordula Artelt, Johannes Hartig, Nina Jude, Olaf Köller, Manfred Prenzel, Wolfgang Schneider & Petra Stanat (eds.), *Pisa 2009: Bilanz nach einem Jahrzehnt*. Münster et al.: Waxmann. 200–230.
- Stivers, Tanja. 2010. An overview of the question–response system in American English conversation. *Journal of Pragmatics* 42. 2772–2781.
- Tomasello, Michael. 2003. *Constructing a language*. Boston: Harvard University Press.
- Voet Cornelli, Barbara, Barbara Geist, Angela Grimm & Petra Schulz. 2011. Wie wird der Sprachstand mehrsprachiger Kinder in pädiatrischen Vorsorgeuntersuchungen erhoben? Erste Ergebnisse aus dem Projekt *cammino*. In Stefan Jeuk & Joachim Schäfer (eds.), *Deutsch als Zweitsprache in Kindertageseinrichtungen und Schulen: Aneignung, Förderung, Unterricht, Beiträge aus dem 7. Workshop Kinder mit Migrationshintergrund*. Freiburg i.Br.: Filibach. 43–73.
- Vygotsky, Lev S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wessel, Lena. 2015. *Fach- und sprachintegrierte Förderung durch Darstellungsvernetzung und Scaffolding. Ein Entwicklungsforschungsprojekt zum Anteilbegriff*. Wiesbaden: Springer Spektrum.
- Wood, David, Jerome S. Bruner & Gail Ross. 1976. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17. 89–100.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge et al.: Cambridge University Press.

