



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/240975/>

Version: Published Version

Article:

Miękisz, A., Cummins, C., Öztürk, Ö et al. (2026) Comparing morphosyntactic and semantic-pragmatic competence in Polish-English bilinguals: toward a language-neutral assessment. *Applied Psycholinguistics*, 47. e23. ISSN: 0142-7164

<https://doi.org/10.1017/s0142716426100575>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ORIGINAL ARTICLE

Comparing morphosyntactic and semantic-pragmatic competence in Polish-English bilinguals: Toward a Language-neutral assessment

Aneta Miękisz¹, Chris Cummins², Özge Öztürk³, Ewa Haman¹ and Napoleon Katsos⁴

¹University of Warsaw, Faculty of Psychology, Poland, ²Linguistics and English Language, University of Edinburgh, UK, ³The University of Sheffield, Faculty of Health, School of Allied Health Professions, Nursing and Midwifery, UK and ⁴Research Centre for English and Applied Linguistics, University of Cambridge, UK

Corresponding author: Aneta Miękisz; Email: amiekisz@psych.uw.edu.pl

(Received 4 September 2025; revised 19 February 2026; accepted 4 March 2026)

Abstract

Evidence that certain semantic and pragmatic skills follow a similar developmental trajectory cross-linguistically suggests an important role for semantic-pragmatic tasks in the assessment of bilingual children. This study investigates whether quantifiers constitute a language-neutral linguistic category in this context. Forty-three Polish–English bilingual children aged 4–7 years completed the Quantifier Comprehension Task (QCT) and the Test for Reception of Grammar (TROG-2), a standardized morphosyntactic assessment, in both languages. The results indicated that children’s performance on QCT is strongly correlated between their two languages. This correlation was significantly stronger than that observed for the TROG-2, indicating that quantifier comprehension may be less language dependent than general grammar comprehension measures. These findings highlight the diagnostic value of semantic-pragmatic tasks, particularly quantifier interpretation, in evaluating bilingual language development. While the children in this study are all typically developing bilinguals, our goal is to explore whether quantifier comprehension tasks can support the development of language-neutral tools for assessing bilingual language skills.

Keywords: bilingualism; developmental language disorder; language assessment; semantic-pragmatic competence; quantifier comprehension

Introduction

In recent years, there has been growing concern over the assessment of bilingual children’s language skills (De Houwer, 2023). This concern is rooted in an increased awareness that the use of tools developed for monolingual populations on bilingual children can result in faulty conclusions (Armon-Lotem et al., 2015; Fleckstein et al.,

2018; Zurer-Pearson, 2010). Bilinguals' language development is somewhat different from that of monolinguals (De Houwer, 1990, 2023; Genesee & Nicoladis, 2007; Kurcz, 2005), as each bilingual child has a unique constellation of relevant linguistic experience when we consider the set of factors including age of onset, input, length of exposure, context of exposure, and language status. The specific features of each language also play an important role in child language development (Byers-Heinlein & Fennell, 2014).

De Houwer (2023) highlights how, as a result of this diversity of experience, bilingual children follow distinct developmental trajectories that cannot be fairly assessed using monolingual norms. In particular, assessments that focus solely on one language, often the societal language, can underestimate a bilingual child's total linguistic capacity and lead to misdiagnosis. Notably, in certain areas of linguistic competence, typically developing bilingual children may score similarly to monolinguals with developmental language disorder (previously also known as specific language impairment (SLI), language impairment, or primary language impairment) (Bishop, 2017; Ebert & Kohnert, 2016; Kohnert et al., 2009).

Developmental language disorder (DLD) is a neurodevelopmental condition affecting children's language skills across domains such as syntax, phonology, semantics, and pragmatics (Bishop, 2017a; Leonard, 2014; Montgomery & Evans, 2009). Symptoms vary cross-linguistically, but when present in bilinguals, DLD typically affects both languages spoken by the child (Armon-Lotem, 2012; Paradis et al., 2003). Diagnosis is usually made between 4 and 7 years through standardized tests of expressive and receptive language, supplemented with caregiver and teacher reports (Bishop et al., 2017b).

For bilingual children, diagnostic challenges are substantial. Because bilinguals follow developmental trajectories that differ from monolingual norms (De Houwer, 2023), using monolingual tests risks both over- and under-identification of DLD (Pert & Bradely, 2018). Some language behaviors typical of bilingual development resemble the profile of DLD (Ebert & Kohnert, 2016; Kohnert et al., 2009), which increases the risk of false positives. Conversely, bilingual children with DLD may go unrecognized if their exposure to the societal language has been insufficient for reliable assessment (Grimm & Schulz, 2014). Indeed, bilingual children are over-represented in referrals for suspected DLD in several clinical settings (e.g., Laasonen et al., 2018) yet may also be underdiagnosed due to limited linguistic experience in the testing language.

These issues highlight the need for diagnostic approaches that take a bilingual child's complete linguistic profile into account (De Houwer, 2023). Although best practice recommends assessing both of a bilingual child's languages (Bedore & Peña, 2008; Gillam et al., 1999; Peña, Bedore & Kester, 2016), this is rarely feasible in practice. For many languages, no standardized tests exist, and obtaining norms for heterogeneous bilingual populations is difficult. Moreover, clinicians may not speak all of the languages they must assess.

These limitations have prompted interest in language-neutral, or minimally language-dependent, assessment tools (Armon-Lotem et al., 2015; Polišenská et al., 2020). Such tools aim to evaluate abilities that develop in similar ways across languages and are therefore less tied to specific morphosyntactic or lexical knowledge. Prominent examples include non-word repetition (e.g., the quasi-universal Q-U

NWRT; Chiat, 2015) and narrative macrostructure tasks, which rely on general cognitive-pragmatic skills that transfer across a bilingual's languages (Gagarina & Lindgren, 2020; Pearson, 2002). Semantic-pragmatic skills, including pronominal reference, quantification, exhaustivity, and scalar implicatures, also appear to follow comparable developmental trajectories across languages (Avrutin, 1999; Katsos et al., 2016; Philip, 1995). Because these skills are less dependent on language-specific morphology, they offer promising candidates for language-neutral assessment.

Building on this work, the present study investigates one such domain: quantifier comprehension. It is important to note that, while this study focuses on typically developing bilingual children, we are motivated by the broader goal of addressing diagnostic challenges associated with DLD, particularly in multilingual populations.

As noted, several factors contribute to the differing pace of linguistic development in a child's languages, including the duration of exposure, language dominance, age of L2 onset, and typological differences. For practical purposes, such as diagnosing DLD, it would be convenient if we were able to assume that shared properties in a bilingual child's languages follow a uniform developmental trajectory. The general developmental milestones have already been identified in early language acquisition, including the universally recognized stages of babbling, one-word utterances, two-word combinations, and multi-word sentences (Berman, 1981; Berk, 2009; Bloom, 1973; Brown, 1973; Gillis & Ravid, 2009; Slobin, 1985, 2001). However, while it is well documented that monolingual and bilingual children reach early key linguistic milestones, ranging from sounds to grammar, within similar timeframes (De Houwer, 2009; Genesee et al., 1995; Genesee et al., 2004; Meisel, 2004; Muszyńska et al., 2025), this cannot be taken as a diagnostic shortcut. Variation between languages can be substantial, especially in morpho-syntactic domains such as tense marking, subject-verb agreement, and negation (Bittner et al., 2011). Therefore, evaluating these skills in only one language does not reliably reflect abilities in another; comprehensive assessment across both languages remains necessary. Moreover, the early milestones, while well-established, are of limited relevance for the assessment of older children.

An alternative approach, advocated by Polišenská et al. (2020), is to develop language-neutral, evidence-based assessments that evaluate fundamental language-learning capacities independent of specific language knowledge. In this context, "language-neutral" features are those that develop similarly across languages and are not specific to the grammar or vocabulary of any one language—that is, those in which a child's performance in one language informs us about the other(s) (Armon-Lotem et al., 2015).

Candidate language-neutral indicators such as non-word processing and narrative tasks have been used as a basis for assessment tools. A striking example is the quasi-universal non-word repetition task (Q-U NWRT) of Chiat (2015). Designed to minimize the influence of knowledge from any specific language, the Q-U NWRT incorporates segmental and phonotactic features common across languages while avoiding language-specific characteristics. The premise of this design is that bilingual children can draw on their overall linguistic competence, making their performance on this task a valid reflection of their underlying language skills (Boerma & Blom, 2017).

Narrative tasks are also commonly used to assess a bilingual child's linguistic abilities (Gagarina & Lindgren, 2020). Specifically, research shows that bilingual children are able to apply knowledge about narrative macrostructure (e.g., story grammar elements such as goals, actions, and outcomes) acquired in their first language (L1) to their second language (L2) (Pearson, 2002). In contrast to narrative microstructure, which encompasses factors such as lexical diversity and relational and referential devices, macrostructure does not appear to be influenced by the amount of language-specific exposure that a child receives in each of their languages and therefore allows linguistic skills in a second or third unassessed language to be inferred from performance in the child's L1 (Boerma & Blom 2017; Hipfner-Boucher *et al.*, 2015; Pearson, 2002; Squires *et al.*, 2014). Recognizing the diagnostic value of such tasks, the LITMUS battery (Language Impairment Testing in Multilingual Settings) developed through COST Action IS0804 (2010–2013) incorporates narrative tasks alongside sentence repetition, nonword repetition, and lexical assessments to support the identification of language impairment in multilingual contexts (Armon-Lotem *et al.*, 2015).

As in the case of the language-neutral skills mentioned above, the acquisition of many semantic and pragmatic properties seems to follow a similar developmental path cross-linguistically. This has been documented for pronominal reference (Avrutin, 1999; Baauw, 2002), quantification (Hanlon, 1987; Katsos *et al.*, 2016; Philip, 1995), exhaustivity in question interpretation (Roeper *et al.*, 2002; Schulz & Roeper, 2011), and scalar implicatures (Noveck, 2001; Papafragou & Musolino, 2003). Therefore, assessing a child's performance with these semantic and pragmatic skills in one language might allow us to approximate their performance in another, unassessed language.

In the current study, we examine one of these semantic categories, namely quantification in Polish-English bilingual children. Specifically, we focus on children's comprehension of quantifying expressions such as "all," "some," "most," and "none" across the two languages. Alongside this, we use the English and Polish versions of the Test for Reception of Grammar (TROG-2) (Bishop, 2003) to assess the same children's understanding of a range of grammatical structures.

Our principal goal in this study is to investigate whether these quantifiers constitute a language-neutral linguistic category in the sense described above: that is, whether a bilingual child's capability in quantifier comprehension in one of their languages can be predicted by their performance when tested in the other language. Under this hypothesis, we expect a strong correlation between children's performance in English and Polish versions of the quantifier comprehension task. In particular, we expect this correlation to be significantly stronger than that between the English and Polish versions of TROG-2, as this tests grammatical categories that are not language-neutral and that in some cases follow different developmental trajectories in the bilingual child's two languages. While this study does not include children with suspected or diagnosed DLD, our motivation stems from the diagnostic challenges DLD poses in bilingual populations. We focus here on typically developing bilingual children in order to evaluate whether quantifier comprehension offers promise as a language-neutral assessment tool that could later be validated in clinical contexts.

In the following sections, we present a more detailed account of the meaning and acquisition of quantifiers before turning to the specifications of the experimental task.

Quantifiers

In natural language, the meaning of the quantifiers such as “all” and “some” is taken to correspond to relations between sets rather than denoting properties of individuals (Barwise & Cooper, 1981). Under this view, the statement “All books are in the boxes” is true if all books under discussion are inside the boxes, and it is false in every other circumstance. The statement “Most of the books are in the boxes” is true if the number of books inside the boxes is greater than the number of books outside and false otherwise. Similarly, the statement ‘Some of the books are in the boxes’ is true if two or more books are in the boxes, and false otherwise.”¹

Quantified sentences have systematic entailment properties. Quantifiers that guarantee inferences from sets to supersets (e.g., “Some students play the piano” entailing “Some students play an instrument”) are known as monotone increasing. Quantifier comprehension involves both semantic and pragmatic processing. While the truth conditions of expressions such as “some” or “none” can be evaluated semantically (e.g., “none” entails an empty intersection between sets), children must also assess whether the utterance is pragmatically appropriate or underinformative. For example, the statement “*some of the toys are in the boxes*” is semantically true even if all of the toys are in the boxes, but it violates expectations of informativeness. Detecting such underinformativeness engages pragmatic reasoning, particularly the ability to interpret utterances relative to Gricean maxims (Grice, 1989; Katsos et al., 2016). The Quantifier Comprehension Task used in this study was therefore designed to elicit both semantically false responses and pragmatically infelicitous responses, enabling a direct assessment of children’s sensitivity to both semantic accuracy and pragmatic appropriateness. In downward-entailing linguistic environments, replacing a general term (i.e., the set) with a more restrictive term (i.e., the subset) preserves truth. Quantifiers that guarantee inferences from sets to subsets (e.g., “None of the students is smoking” entailing “None of the students is smoking expensive cigars”) are known as monotone decreasing.

Many human languages contain quantifiers, and these quantifiers have similar entailment properties cross-linguistically (Von Stechow & Matthews, 2008). They also appear to impose similar usage constraints, such as the need to be informative. For instance, speakers should not describe a situation in which all students are running by saying, “Some students are running.” Although strictly speaking this statement is true, it is an underinformative statement, and it potentially invites the listener to draw the further conversational inference that the speaker has reason not to utter the more informative statement “All students are running” (Grice, 1989; Horn, 1972).

Acquisition of quantifiers

Young children’s comprehension of quantified expressions has received significant attention in the literature (Crain, 2017 i.a.; Gualmini, 2003; Katsos et al., 2016; Musolino, 1998; Papafragou & Musolino, 2003; Pouscoulous et al., 2007).

In general, children appear to have mastered several of the semantic properties of quantifiers at around age six, but their responses to sentences with quantifiers differ from adults': children are more likely than adults to accept underinformative quantified expressions (Noveck, 2001). Several accounts of this phenomenon have been offered in previous literature, including appeal to children's immature cognitive capacities (Inhelder & Piaget, 1964 i.a.) or immature grammars (i.e., difficulties in mapping form to meaning in quantifier comprehension; Philip, 1995). However, researchers have yet to reach consensus on the source of young children's non-adult-like responses to quantified expressions (Crain, 2017).

Katsos *et al.* (2016) examined children's understanding of quantified expressions across 31 languages from 11 families. The results revealed a cross-linguistically similar order of acquisition that accorded with four main constraints posited by the authors: monotonicity, totality, complexity, and informativeness. Specifically, in 27 of the 31 languages tested, children performed better with monotone increasing quantifiers ("all," "some") than with their monotone decreasing counterparts ("none," "some... not"). In 25 of the 31 languages tested, children performed better with total than partial quantifiers—that is, with "all" and "none," which attribute the same property to every member of the set, compared to "some" and "some... not," which do not. Children were more successful at comprehending "some" than "most" in all 31 languages, which the authors take as evidence for the role of complexity: they argue that, in order to successfully understand "Most of the As are Bs," children need to be able to compare the cardinalities of the set of As that are Bs with those of the set of As that are not Bs. By contrast, "Some As are Bs" is less complex because children do not need to restrict the quantifier to a specific set of entities or compare cardinalities in order to evaluate the existential claim that is being made. Finally, the results also confirm that in all 31 languages studied, children follow the "informativeness" constraint: they are aware of violations of pragmatic felicity arising from under-informativeness, although they treat these violations less strictly than violations of truth. These findings broadly chime with proposals arguing for cross-linguistic similarities in the meaning and use of quantifiers. Slim *et al.* (2021) provided evidence from adult bilinguals that quantifiers are not only interpreted similarly across languages but also share their underlying logical form. In sentence-picture matching tasks, Dutch–French bilinguals showed cross-linguistic priming of quantificational scope, suggesting that the logical representations constructed for quantifiers in one language are accessible and used when processing the other language.

Drawing upon these findings, in this study, we explore the usefulness of bilingual children's performance on a quantifier interpretation task in one language as a predictor of their performance on that task in the other language. We compare this to the cross-linguistic predictive usefulness of their performance on a task designed to measure complex linguistic skills, which are typically strongly language-specific. The ultimate goal of this approach is to exploit the similarities in the developmental trajectories of quantifiers across languages by using quantifier interpretation in bilingual assessment.

In attempting this, we take the view that, although performance with quantifiers requires some measure of numerical skill (Pietroski *et al.*, 2009; Shikhare *et al.*, 2015), it is more closely associated with linguistic abilities. This is because quantifier comprehension depends not just on estimating quantity but also on understanding

set-theoretic relations, logical entailments, and pragmatic appropriateness, features that are encoded linguistically and governed by semantic and pragmatic rules (Cheng et al., 2013; Dolscheid & Penke, 2018; Katsos et al., 2012). For example, interpreting “some” as pragmatically distinct from “all” requires knowledge of scalar implicatures, not just quantity. We consider this study a first step towards directing attention to the possibility of using quantifiers as part of the assessment of bilingual children’s linguistic aptitude.

General methods

Participants

A total of 45 Polish-English bilingual children living in the east of England participated (mean age: 6;10, range: 4;11–7;10). Mean age of onset of English was 2;3 (range: birth–4;1). The mean length of exposure to English was 3;10 (range: 1;8–6;2). Participants did not have a history of autism or hearing or neurological impairment. Both of their parents were Polish: around 30% of the parents had college education, with around 60% reporting secondary and around 10% elementary education. In the final analysis, 43 children were included (22 males and 21 females). One child was excluded due to strong acquiescence bias (responding “Yes” to all items), and one child was excluded for failing the pre-test for the quantifier task (see below).

Measures and procedures

The following tasks were used: (1) the Quantifier Comprehension Task in English and in Polish (Katsos et al., 2012); (2) the Test for Reception of Grammar (TROG-2) (Bishop, 2003) in English and in Polish; and (3) the Non-verbal IQ test (Raven, 2003). Children were tested individually at nurseries or primary schools, in a quiet room outside their classroom, following the ethical protocols designated by the host institution, the University of Cambridge. Parents were asked to sign a Parental Consent Form (provided in two language versions: Polish and English) and to fill in a short questionnaire collecting data about their child’s onset of exposure to English and length of exposure, measured as length of time in the UK. This included information about the language used at home, the language used with the child, the parent’s level of education, their profession, and the child’s birth order and number of siblings. The parents were asked if they had any concerns about their child’s development and confirmed that their child had no history of autism and no hearing or neurological impairment. The experimenters also made sure that the children independently agreed to take part in all of the activities involved in collecting the data.

The tasks were administered in a counterbalanced order both at the level of language (Polish-English, English-Polish) and within language (TROG-Quantifiers, Quantifiers-TROG), except for the non-verbal IQ test, which was administered in a Polish adaptation only (Raven, 2003) and which was always used as the last test in the session.

Measures: Quantifier Comprehension Task

In the Quantifier Comprehension Task (for details, see Katsos et al., 2012), participants were seated in front of a laptop computer and were shown visual displays

on the screen while listening to prerecorded sentences. To ensure cross-linguistic parity, all QCT stimuli were pre-recorded by a female native speaker following a matched recording script and consistent prosodic parameters across both language versions. Furthermore, the Polish version was not a direct translation but was developed collaboratively within COST Action A33 as part of a coordinated, cross-linguistic empirical project. As documented in Katsos *et al.* (2012), all participating language teams followed shared theoretical principles and common item structures and agreed on methodological constraints when selecting and validating quantifier expressions, ensuring that all language versions, including Polish, were prepared in parallel and under comparable design assumptions.

Before the experiment started, the experimenter introduced the participant to the Cavegirl, a fictional cartoon character who appears on the screen. The participant was told that her task was to help the Cavegirl learn better the language (English or Polish). The Cavegirl would say how many toys there are in the boxes. If what she said was right, the participant should tell her, "That was right." If what she said was wrong, the participant should tell her, "That was wrong," and, in order to help her learn, they should tell her why it was wrong. The experimental task was preceded by a warm-up session where children were familiarized with the Cavegirl, the task demands, and the pictures of the objects mentioned in the sentences. Children were shown a picture of each of the objects on the computer screen. The experimenter then pointed to each object and named it. The objects were chosen from the domain of household items, food, musical instruments, and toys, and teachers independently confirmed that the participants knew these objects.

After participants were familiarized with the task, their competence with numbers was tested with five statements, one for each number from one to five, three of which were true and two of which were false. This was done to confirm that participants could count up to five in order to ensure that children could make correct judgments about quantity when simple counting is involved. A discontinue rule was applied, under which, if the child failed an item in the warm-up phase, the child was asked to be a bit more careful, and the sentence was repeated. If a participant failed in the repetition of an item, or if he/she failed in the first attempt in three or more warm-ups, then the experiment was stopped. In this study, one child was excluded at this stage.

In each experimental trial, the Cavegirl produced a single utterance of the type "[Quantifier] of the [objects] are (not) in the boxes." As instructed at the beginning, children evaluated whether what the Cavegirl said was "right" or "wrong" and, if they said "wrong," were encouraged to justify why they did so. This was done in order to check whether the participant was rejecting a statement for reasons unrelated to falsity or informativeness. In this study, all justifications of rejections, whether correct or incorrect, mentioned a quantity-related word or deictic expression often combined with a spatial expression (e.g., "Because these are out"), which suggests that children responded based on the appropriateness of the quantifier rather than some other aspect of the sentence.

Seven quantifying expressions were tested: "some," "most," "some . . . not," "all," "none," "all . . . not," and "not all." For each quantifier tested, two types of visual situations were used, one which rendered the utterance true and informative and one which rendered the utterance false. For "some," "most," and "some . . . not," a

third type of situation was also used, which rendered the utterance true but pragmatically underinformative. In the case of “some” and “most,” this was a display in which all of the objects were in the boxes; in the case of “some . . . not,” it was a display in which none of the objects were in the boxes.

Six blocks of items were created, with the order of items pseudo-randomized within each block to avoid mention of the same quantifier or the same object in adjacent items. The blocks were presented in one of three different orders.

Measures: Test for Reception of Grammar (TROG-2)

Participants were assessed using TROG-2 (Bishop, 2003) and its Polish translation (Smoczyńska, 2008, unpublished manuscript submitted to Pearson Assessment) as a measure of receptive grammar. TROG is an individually administered, standardized test of morphosyntax. It has been used to assess children with specific language disorders, deafness, intellectual disability, and cerebral palsy. The test is appropriate for children aged 4 to 13 years and has been normed with more than 2000 British children who did not have any known learning difficulty, hearing loss, or other disability. The results of the standardization were found to be valid even after controlling for social background (Bishop, 1989).

Each test item is presented in a multiple-choice format with four pictures presented on a single board. One of the pictures illustrates the target structure, and three constitute the lexical and grammatical foils to this structure. The child is auditorily presented with the stimulus, containing a particular grammatical structure. Then the experimenter asks the child to point to one of four pictures that best corresponds to what he/she has heard.

The test consists of 20 blocks of 4 items. Each block assesses the child's comprehension of a specific type of morphosyntactic construct (e.g., singular/plural personal pronouns, masculine/feminine personal pronouns, nouns, comparatives, verbs, reversible passives, negation, prepositional phrases, relative clauses), arranged in increasing order of difficulty (as established for the English version). The test is usually scored according to the number of blocks successfully passed (where passing a block involves the child responding correctly to all 4 items, the probability of doing so by chance being 0.004). TROG is designed to use a restricted and simple set of lexical items (e.g., elephant, girl, boy, dog, table, etc.) in order to minimize the likelihood of the participant's failure due to lack of knowledge of individual words.

The Polish version of TROG is an official translation accepted by Pearson Co., although so far it lacks psychometric properties. However, there are indications that it closely resembles the original English version in terms of results: in a study of Polish-English bilinguals compared to Polish and English monolinguals (age 7 to 10 years), average TROG results for all three groups (total $n = 78$) in either Polish or English were very close to each other (Grose-Hodge et al., 2024).

Non-verbal IQ

Non-verbal IQ was tested with Raven's Coloured Progressive Matrices (CPM, Raven 2003). CPM is one of the most well-known, most widely researched, and most

widely used of all culture-reduced tests. The test demonstrates high reliability and validity across a wide range of populations (Raven *et al.*, 1996). Although the CPM is a non-verbal test, national standardizations are essential to ensure that administration procedures, instructions, and normative comparisons are culturally and linguistically appropriate for the target population. The version used in this study was the Polish standardization, which includes norms and protocols validated for Polish-speaking children (Raven, 2003; Szustrowa & Jaworowska, 2003).

Results

In the Quantifier Comprehension Task administered in Polish, participants were 82.0% accurate on semantically determined conditions (those which were true and informative or false), with individual performance ranging from 47.2% to 100% (SD 16.5%) and 55.6% accurate on pragmatically determined conditions (that is, they rejected underinformative utterances at this rate), with individual performance ranging from 0% to 100% (SD 30.2%). The overall accuracy (weighting semantic and pragmatic scores equally) was 68.8% (range 27.8%–98.6%, SD 18.7%). In the Quantifier Comprehension Task administered in English, the corresponding figures were 75.5% (range 38.9%–100%, SD 15.7%) for the semantic conditions and 57.9% (range 0%–100%, SD 28.5%) for the pragmatic conditions, with an overall accuracy of 66.7% (range 32.6%–97.2%, SD 17.0%).

We do not assume that the outcome measures are normally distributed², but in what follows we report parametric statistics on the basis of the sufficiently large sample size ($n = 43$). For completeness, we include the corresponding non-parametric test results in footnotes. Paired *t*-tests disclose that the participants were significantly more accurate on the semantic conditions in the Polish task than in the English task ($t = 3.93$, $p < 0.001$), but there was no significant difference between languages in the pragmatic conditions ($t = 0.74$, $p = 0.47$) or on the overall measure ($t = 1.37$, $p = 0.18$).³

In the TROG-2 administered in Polish, participants passed an average of 10.1 blocks (range 3–19, SD 4.42) and got an average of 53.2 items correct (range 22–77, SD 17.34). In the English version of TROG-2, participants passed an average of 6.88 blocks (range 1–16, SD 3.21) and got an average of 37.9 items correct (range 5–69, SD 14.54). Paired *t*-tests show that this represents significantly higher performance in the Polish than the English version of TROG-2 (on the item measure, $t = 5.34$; on the block measure, $t = 4.94$; both $p < 0.001$).⁴

Participants' results on the CPM were age-appropriate. Specifically, their average score was 21.4 (range 14–32, SD 4.87). The corresponding *sten*⁵ scores ranged from 4 to 10, with a mean of 6.91 (SD = 1.59).

Correlations between measures⁶

The correlation between the English and Polish QCT scores was 0.776 for the semantic conditions and 0.772 for the pragmatic conditions. The correlation between the English and Polish overall QCT scores was 0.843. The correlation between the English and Polish TROG scores, as measured by the number of blocks passed, was 0.422 (and as measured by the number of items passed, 0.310).⁷

Comparing the semantic QCT correlation with the TROG (by block) correlation, we find a significant difference ($Z = 2.62$, $p = 0.009$); comparing the pragmatic QCT correlation with the TROG (by block) correlation in the same way, we again find a significant difference ($Z = 2.57$, $p = 0.010$); and the same is true for the composite QCT measure ($Z = 3.49$, $p < 0.001$).⁸

The correlation between the English and Polish TROG raw scores is significantly lower than the English and Polish QCT correlations for the semantic condition ($Z = 3.20$, $p = 0.0014$), the pragmatic condition ($Z = 3.16$, $p = 0.0016$), and the composite QCT measure ($Z = 4.07$, $p < 0.001$).⁹

Therefore, our main expectation of a strong correlation between children's performance in English and Polish versions of the Quantifier Comprehension Task was confirmed. In particular, we expected this correlation to be significantly stronger than that between the English and Polish versions of TROG-2, and our results confirmed this hypothesis.

Polish TROG scores by block were moderately correlated with the English QCT semantic scores ($r = 0.570$), weakly correlated with the English QCT pragmatic scores ($r = 0.207$), and moderately correlated with the English QCT overall scores ($r = 0.437$).¹⁰ The same is true if we consider Polish TROG raw scores ($r = 0.535$, $r = 0.265$, and $r = 0.469$, respectively).¹¹ Similarly, English TROG scores by block were moderately correlated with the Polish QCT semantic scores ($r = 0.562$), weakly correlated with the Polish QCT pragmatic scores ($r = 0.268$), and moderately correlated with the Polish QCT overall scores ($r = 0.465$).¹² Again, the same is true if we consider English TROG raw scores ($r = 0.533$, $r = 0.214$, and $r = 0.408$, respectively).¹³

Finally, CPM scores were moderately correlated with the Polish TROG scores by block ($r = 0.434$) and in raw score ($r = 0.389$); English TROG scores by block ($r = 0.505$) and in raw score ($r = 0.469$); Polish semantic QCT scores ($r = 0.359$); Polish overall QCT scores ($r = 0.328$); English semantic QCT scores ($r = 0.397$); and English overall QCT scores ($r = 0.402$). They were weakly correlated with the Polish pragmatic QCT scores ($r = 0.210$) and English pragmatic QCT scores ($r = 0.262$).¹⁴

Predicting scores

Thinking of these measures as potential predictors of the TROG scores in the other language, we find that the Polish semantic QCT is a numerically better predictor of English TROG scores than are either the Polish TROG or the CPM, and the English semantic QCT is a numerically better predictor of Polish TROG scores than are either the English TROG scores or the CPM. The other-language pragmatic QCT measures are numerically worse predictors than the other-language TROG scores or the CPM. This applies whether we consider the TROG block score or raw score.

Discussion

Bilingualism presents a challenge to language assessment because children's skills may differ across their two languages. There are no clear norms that reflect the diversity of bilingual development, and bilingual professionals and/or suitable tests

are often unavailable for both languages. As a consequence, bilingual children are typically evaluated in the dominant language of their community and with reference to monolingual norms, with adverse implications for the accuracy of the diagnosis of DLD. One prospective solution to this problem is an increased focus on language-neutral skills, the aspiration being to develop a test battery that enables professionals to estimate language skills in an unassessed language based on a child's skills as assessed in another language (Pearson, 2002; Squires *et al.*, 2014).

Our work explored one possible contributor to such a test battery. Specifically, drawing on research suggesting that certain semantic and pragmatic skills follow a similar developmental trajectory cross-linguistically (Slobin, 1985; Thordardottir, 2015), we wished to investigate whether we could support the stronger claim that these skills are language-neutral in the sense required for assessment. To that end, we examined Polish- and English-speaking bilingual children's comprehension of quantifier expressions such as "all," "some," "most," and "none" across these two languages and tested whether we can use the bilingual child's quantifier comprehension performance in one language as an approximator of their corresponding skills in the other language. We compared the children's performance on the Quantifier Comprehension Task (QCT) in Polish and English, and their performance on a well-studied morphosyntactic evaluation task, the Test for Reception of Grammar (TROG-2), in both languages.

We hypothesized that, since quantifier comprehension may be a language-neutral skill, the results of the QCT would be more highly correlated between languages than the results of TROG. Our results confirmed this hypothesis, indicating that quantifier comprehension abilities are more consistent across a bilingual child's languages than are grammatical abilities in general.

These results have important implications for bilingual language assessment. As noted above, speech and language therapists (SLTs) and other language professionals are frequently tasked with evaluating bilingual children's language skills. However, given the diversity of languages involved, it is impossible for an SLT to be able to speak all the languages spoken by the children they need to evaluate. Hence, the task of finding suitable proxy measures for a child's linguistic skills in their other languages is a pressing one. Our results are promising in that they demonstrate the QCT to be a relatively stable test instrument across languages for bilingual children, compared to standard morphosyntactic measures.

The question then arises of whether, and to what extent, the QCT is tapping general linguistic abilities. Could it be that the QCT results are instead a reflection of non-linguistic cognitive capabilities, such as those for analogical reasoning? There are reasons to doubt this interpretation. Firstly, as demonstrated in this study, the QCT results are not strongly predicted by the non-verbal IQ measure used, Raven's Coloured Progressive Matrices (CPM); indeed, there was no significant difference in correlation between CPM and QCT scores and between CPM and TROG scores.

Secondly, previous research by Katsos *et al.* (2011) suggests that the QCT is associated with language skills in monolinguals. Specifically, Katsos *et al.* (2011) compared Spanish-speaking children with diagnoses of Specific Language Impairment (SLI, currently labeled DLD) ($n = 29$, mean age: 6;6, age range 4;0–6;9, 20 male) with age-matched and language-matched typically developing controls. Recall that DLD is characterized by difficulty with language that is not caused by a

known neurological, sensory, intellectual, or emotional deficit, and consequently, non-linguistic cognition in DLD is generally expected to be age-appropriate.

Katsos et al. (2011) evaluated their participants' non-verbal IQ by the Raven's Standard Progressive Matrices (SPM) test and the Gestalt closure subtest of the Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004), and results for the DLD group were age-appropriate. In another study, participants were also tested on the Spanish adaptation of the QCT (Katsos & Smith, 2010). Here, the participants with DLD scored significantly lower than their age-matched controls, but their results did not significantly differ from the younger language-matched controls. This suggests that the QCT is indeed tapping linguistic rather than non-linguistic cognition.

Our findings may contribute to the need identified by Polišenská et al. (2020), who advocate for the use of language-neutral assessment tools that target core language-learning abilities rather than language-specific knowledge. Their recommendations for tools such as the Cross-linguistic Nonword Repetition (CL-NWR) task, the Early Sociocognitive Battery (ESB), and dynamic assessment strategies provide concrete avenues for addressing the diagnostic challenges for DLD posed by multilingualism. Each of these tools is designed to be culturally inclusive and minimally biased by specific language exposure, thereby enabling more equitable and accurate identification of language disorders in multilingual children. The QCT, as demonstrated in this study, aligns conceptually with these proposals by offering a method to evaluate a bilingual child's underlying language competence in a way that is robust across languages. Integrating such tools into a broader assessment battery may form the foundation for more inclusive, scalable, and reliable diagnostic practices in multilingual contexts. However, it remains to be established precisely which linguistic abilities are tapped by the QCT and how it might best fit into a battery of tasks designed to evaluate bilingual children's language abilities through the medium of a single language.

One of the limitations of the current study is the relatively small set of quantificational expressions included in the QCT. Although these seven expressions ("some," "most," "none," etc.) were chosen to capture key logical and pragmatic contrasts, they represent only a subset of the broader domain of pragmatic competence. To build a more comprehensive profile of bilingual children's pragmatic skills, future research should examine areas such as scalar implicature beyond quantifiers (e.g., "or" and "might") and referential communication (e.g., use of definite vs. indefinite expressions). These pragmatic abilities also develop early and could help to extend and refine the current findings.

Some caution is also advisable regarding the characteristics of the sample, as the final sample ($N = 43$) reflected the size of the accessible population and practical recruitment constraints in local schools. Although geographically concentrated in the East of England, the participating children came from first-generation Polish immigrant families originating from diverse regions across Poland. Nevertheless, the sample is not fully representative of all bilingual profiles, and this limits the generalizability of the findings. Future research should therefore extend this paradigm to larger samples and to bilingual populations with different linguistic and sociolinguistic backgrounds.

In sum, our work points to the potential value of the Quantifier Comprehension Task in the assessment of bilingual children. Our results indicated that children's performance on this task is strongly correlated between their two languages. This opens up the possibility of treating quantification as a potentially language-neutral linguistic category, which might be productively used in the assessment of bilingual children, in that the results of a test of this kind administered in one language are predictive of the results of a corresponding test in the other language. Given the pressing need for better assessment of language development in bilinguals (Peña *et al.*, 2020), we argue that our results should motivate further investigation into the use and development of tools that include quantifier comprehension as a component of language assessment.

Acknowledgements. We are grateful to the anonymous reviewers for their constructive feedback, including the suggestion to conduct additional analyses. This research was supported by the European Cooperation in Science and Technology (COST) Actions A33 *Cross-Linguistically Robust Stages of Children's Linguistic Performance* and IS0804 *Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*, the British Academy, and the Polish Ministry of Science and Higher Education. Data collection was partly funded by Grant no. 809/N-COST/2010/0 from the Polish Ministry of Science and Higher Education/National Science Centre, *Rozwój poznawczy i językowy polskich dzieci dwujęzycznych u progu edukacji szkolnej—szanse i zagrożenia* [Cognitive and Linguistic Development of Polish Bilingual Children: Risks and Opportunities at the School Entrance Age]. Additional support was provided by COST Action IS0804 and A33 STMS (Short Term Scientific Mission) grants; the British Academy (SG-090676) *Bilingual language development: a comparison of morphosyntactic and semantic-pragmatic competence*; and EuroXPRAG grants.

Notes

- 1 This is arguable for the case in which exactly one book is in the boxes, as the extent to which 'some' encodes plurality is not entirely clear. However, this is not relevant for the materials used in this study.
- 2 Shapiro-Wilk tests disclose a significant deviation from normality ($p < 0.05$) for the following variables: Polish semantic QCT score, Polish pragmatic QCT score, English semantic QCT score, English overall QCT score, Polish TROG-2 blocks, Polish TROG-2 items correct, CPM score, and CPM sten score.
- 3 Wilcoxon signed-rank tests similarly disclosed a significant difference on the semantic conditions ($V = 665$, $p < 0.001$) but no significant difference in the pragmatic conditions ($V = 257.5$, $p = 0.35$) or the overall measure ($V = 525$, $p = 0.22$).
- 4 Wilcoxon signed-rank tests also disclosed significant differences (on the item measure, $V = 825$, $p < 0.001$; on the block measure, $V = 765$, $p < 0.001$).
- 5 Sten scores (short for "standard ten") are normalized scores on a 10-point scale, with a mean of 5.5 and standard deviation of 2, where scores of 1–3 are considered low, 4–6 average, and 7–10 high.
- 6 Additionally, we examined whether age of onset (AoA) or length of exposure to English were associated with cross-language performance. Zero-order correlations between AoA/exposure and task scores were small (mostly $|r| < .20$), with slightly higher values for TROG than for QCT. Splitting the sample at the median AoA or exposure did not change the overall pattern: QCT correlations remained uniformly high across groups (0.81–0.89), whereas TROG correlations showed greater variability (0.14–0.57), though subgroup sizes were small. These additional analyses indicate that AoA and exposure do not substantially affect the observed cross-linguistic consistency, particularly for QCT.
- 7 The corresponding Spearman's rho correlation values are 0.734 for the semantic QCT, 0.749 for the pragmatic QCT, 0.835 for the overall QCT, 0.368 for the TROG scores by block, and 0.316 for the TROG scores by item.
- 8 This also holds for the Spearman's rho values, applying the same test: $Z = 2.47$ ($p = 0.014$), $Z = 2.61$ ($p = 0.0091$), and $Z = 3.66$ ($p < 0.001$), respectively.

- 9 This also holds for the Spearman's rho values: $Z = 2.73$ ($p = 0.0063$), $Z = 2.88$ ($p = 0.0040$) and $Z = 3.92$ ($p < 0.001$) respectively.
- 10 The corresponding Spearman's rho values are 0.557, 0.225 and 0.439.
- 11 The corresponding Spearman's rho values are 0.553, 0.237 and 0.462.
- 12 The corresponding Spearman's rho values are 0.573, 0.334 and 0.493.
- 13 The corresponding Spearman's rho values are 0.512, 0.296 and 0.437.
- 14 The corresponding Spearman's rho values are 0.488, 0.464, 0.500, 0.467, 0.370, 0.327, 0.376, 0.419, 0.228, and 0.265. The interpretation as moderate and weak correlations, respectively, remains the same.

References

- Armon-Lotem, S. (2012). Specific language impairment in bilingual children. *Bilingualism: Language and Cognition*, 15(1), 1–4.
- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Multilingual Matters. <https://doi.org/10.21832/9781783093137>
- Avrutin, S. (1999). *Development of the syntax-discourse interface* (Vol. 23). Springer Science & Business Media.
- Baauw, S. (2002). *Grammatical features and the acquisition of reference. A comparative study of Dutch and Spanish. Outstanding dissertations in linguistics*. Routledge.
- Barwise, J., & Cooper, R. (1981). Generalized Quantifiers and Natural Language. In J. Barwise, H. J. Etchemendy, & J. Perry (Eds.), *Philosophy, language, and artificial intelligence* (pp. 241–301). Springer.
- Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, 11(1), 1–29.
- Berk, L. E. (2009). *Child development* (8th Edn.). Pearson
- Berman, R. A. (1981). Regularity vs anomaly: the acquisition of Hebrew inflectional morphology. *Journal of Child Language*, 8(02), 265–282
- Bishop, D. V. (2017a). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, 52(6), 671–680.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Catalise-2 Consortium. (2017b). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080.
- Bishop, D. V. M. (1989). *Test for the reception of grammar (TROG)*. Age and Cognitive Performance Research Centre.
- Bishop, D. V. M. (2003). *Test for reception of grammar: TROG-2 manual*. Pearson.
- Bittner, D., Dressler, W. U., & Kilani-Schoch, M. (Eds.). (2011). *Development of verb inflection in first language acquisition: A cross-linguistic perspective* (Vol. 21). Walter de Gruyter.
- Bloom, L. (1973). *One word at a time: the use of single-word utterances before syntax*. Mouton.
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible?. *Journal of Communication Disorders*, 66, 65–76.
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology*, 56(2), 274–291.
- Cheng, D., Zhou, A., Yu, X., Chen, C., Jia, J., & Zhou, X. (2013). Quantifier processing can be dissociated from numerical processing: Evidence from semantic dementia patients. *Neuropsychologia*, 51(11), 2172–2183.
- Chiat, S. (2015). Nonword Repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (Communication Disorders Across Languages, Vol. 13, pp. 125–150). Multilingual Matters.
- Crain, S. (2017). Acquisition of quantifiers. *Annual Review of Linguistics*, 3, 219–243.
- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press.
- De Houwer, A. (2009). *Bilingual first language acquisition*. Multilingual Matters.

- De Houwer, A. (2023). The danger of bilingual–monolingual comparisons in applied psycholinguistic research. *Applied Psycholinguistics*, 44(3), 343–357.
- Dolscheid, S., & Penke, M. (2018). Quantifier comprehension is linked to linguistic rather than to numerical skills. Evidence from children with Down syndrome and Williams syndrome. *PloS one*, 13(6), e0199743.
- Ebert, K. D., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language Teaching*, 49(03), 301–338.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., & Zebib, R. (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 25(1), 85–101.
- Gagarina, N., & Lindgren, J. (2020). Preface: New language versions of MAIN: Multilingual assessment instrument for narratives–revised. *ZAS Papers in Linguistics*, 64, ix–xii.
- Genesee, F., & Nicoladis, E. (2007). *Bilingual acquisition*. In *handbook of language development*, (E. Hoff, & M. Shatz, Eds.) (pp. 324–342). Blackwell.
- Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, 22, 611–632.
- Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development & disorders: A handbook on bilingualism & second language learning*, (vol 11). Paul H Brookes Publishing.
- Gillam, R. B., Peña, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders*, 20(1), 33–47.
- Gillis, S., & Ravid, D. (2009). Language Acquisition. In: D. Sandra, J. O. Östman, & J. Verschueren (Eds.), *Cognition and pragmatics* (Vol. 3, pp. 201–249). John Benjamins Publishing.
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
- Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research*, 7(4), 821–841.
- Grose-Hodge, M., Dabrowska, E., & Divjak, D. (2024). Beyond accuracy: Fluency and complexity in the spoken output of bilingual and monolingual preadolescents'. *Journal of Monolingual and Bilingual Speech*, 5(3), 328–356. <https://doi.org/10.1558/jmbs.23513>
- Gualmini, A., Meroni, L., & Crain, S. (2003). An asymmetric universal in child language. *Proceedings of Sinn und Bedeutung*, 7, 136–148.
- Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2, 16.
- Hanlon, C. C. (1987). Acquisition of set-relational quantifiers in early childhood. *Genetic, Social, and General Psychology Monographs*, 113(2), 213–264.
- Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, 19(6), 677–692.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. (EA Lunzer & D. Papert, Trans.). Routledge & Kegan Paul.
- Katsos, N., Cummins, C., Ezeizabarrena, M. J., Gavarró, A., Kraljević, J. K., Hrzica, G., Grohmann, K. K., Skordi, A., Jensen de López, K., Sundahl, L., van Hout, A., Hollebrandse, B., Overweg, J., Faber, M., . . . Noveck, I. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, 113(33), 9244–9249. <https://doi.org/10.1073/pnas.1601341113>
- Katsos, N., Ezeizabarrena, M.-J., Gavarró, A., Kuvač Kraljević, J., Hrzica, G., Grohmann, K., Skordi, A., Jensen de López, K., Sundahl, L., van Hout, A., Hollebrandse, B., Overweg, J., Faber, M., van Koert, M., Cummins, C., Smith, N., Vija, M., Parm, S., Kunnari, S., . . . Noveck, I. (2012). The acquisition of quantification across languages: Some predictions. In A. K. Biller, E. Y. Chung, & A. Kimball (Eds.), *Proceedings of the 26th Boston University Conference on Language Development* (pp. 258–268). Cascadia Press.
- Katsos, N., Roqueta, C. A., Estevan, R. A. C., & Cummins, C. (2011). Are children with specific language impairment competent with the pragmatics and logic of quantification? *Cognition*, 119(1), 43–57.
- Katsos, N., & Smith, N. (2010). Pragmatic Tolerance or a Speaker–Comprehender Asymmetry in the Acquisition of Informativeness. In *Proceedings of the 34th Annual Boston Conference in Language Development*, Cascadia Press, MA, USA.

- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children*, 2nd Edn. (KABC-II). AGS Publishing.
- Kohnert K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal Communication Disorders*, 43(6), 456–473.
- Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or “specific” language impairment and children learning a second language. *Brain and Language*, 109(2–3), 101–111.
- Kurcz, I. (2005). *Psychologia języka i komunikacji*. Scholar, in Polish.
- Laasonen, M., Smolander, S., Lahti-Nuuttila, P., Leminen, M., Lajunen, H. R., Heinonen, K., Pesonen, A. K., Bailey, T. M., Pothos, E. M., Kujala, T., Leppänen, P., Bartlett, C. W., Geneid, A., Lauronen, L., Service, E., Kunnari, S., & Arkkila, E. (2018). Understanding developmental language disorder – the Helsinki longitudinal SLI study (HelSLI): a study protocol. *BMC psychology*, 6(1), 24.
- Leonard L. B. (2014). Specific language impairment across languages. *Child development perspectives*, 8(1), 1–5. <https://doi.org/10.1111/cdep.12053>
- Meisel, J. (2004). *The bilingual child. w: The handbook of bilingualism*. (T. Bhatia, & W. Ritchie, Ed.) (pp. 91–113). Blackwell.
- Montgomery, J.W. and Evans, J.L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 269–288.
- Musolino, J. (1998). *Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier negation interaction in English*. University of Maryland.
- Muszynska, K., Krajewski, G., Dynak, A., Garmann, N. G., Romoren, A. S. H., Łuniewska, M., Alcock, K., Katsos, N., Simonsen, H.G., Hansen, P., Krysztofiak, M., Sobota, K., & Haman, E. (2025). Bilingual children reach early language milestones at the same age as monolingual peers. *Journal of Child Language*, 1–24.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165–188.
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, 86(3), 253–282.
- Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). Bilingual children with specific language impairment: How do they compare with their monolingual peers. *Journal of Speech, Language and Hearing Research*, 46, 1–15.
- Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning*, 2nd Edn. Brookes.
- Pearson, B. Z. (2002). Narrative Competence Among Monolingual and Bilingual School Children in Miami. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Multilingual Matters. <https://doi.org/10.21832/9781853595721-00>
- Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192–202
- Peña, E. D., Bedore, L. M., Lugo-Neris, M. J., & Albudoor, N. (2020). Identifying developmental language disorder in school age bilinguals: Semantics, grammar, and narratives. *Language Assessment Quarterly*, 17(5), 541–558.
- Pert, S., & Bradley, B. (2018). *Clinical guidelines for speech and language therapists: Bilingualism: Working with bilingual clients/patients with speech, language and communication needs*. Royal College of Speech and Language Therapists. https://chrome-extension://efaidnbmnnnbpcjpcglclefindmkaj/https://pure.manchester.ac.uk/ws/portalfiles/portal/69559895/mainbilingualism_1_.pdf
- Philip, W. (1995) Event Quantification in the Acquisition of Universal Quantification, PhD thesis, University of Massachusetts, Amherst.
- Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of ‘most’: Semantics, numerosity and psychology. *Mind & Language*, 24(5), 554–585.
- Pinker, S. (1984). *Language learnability and language development*. Harvard University Press.
- Polišenská, K., Chiat, S., Fenton, J., & Roy, P. (2020). Assessing Young Children from Diverse Backgrounds: Novel Ways to Measure Language Abilities and Meet the Requirements of the Early Years Foundation Stage. *Languages, Society & Policy*. <https://doi.org/10.17863/CAM.54064>

- Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicative production. *Language acquisition*, 14(4), 347–375.
- Raven, J. C. (2003). *Raven's coloured progressive matrices – Coloured version* (T. Szustrowa & A. Jaworowska, Trans.). Pracownia Testów Psychologicznych PTP. (Polish standardisation edition)
- Raven, J. C. (2003). *Raven's Coloured Progressive Matrices (CPM)*. Pearson.
- Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's progressive matrices and vocabulary scales: Section 3. Standard progressive matrices*. Oxford Psychologists Press.
- Raven, J. C., JH Court, & Raven, J. (1996). *Raven manual: Section 3 standard progressive matrices with adult US norms by JC Raven, JH Court and J. Raven*. Oxford Psychologist Press.
- Roeper, T., Pearson, B. Z., Penner, Z., & Schultz, P. (2002, July). The Emergence of Wh-Variables: Cross-linguistic Explorations. IX International Congress for the Study of Child Language (IASCL) and the Symposium on Research in Child Language Disorders (SRCLD), University of Wisconsin.
- Schulz, P., & Roeper, T. (2011). Acquisition of exhaustivity in wh-questions: A semantic dimension of SLI?. *Lingua*, 121(3), 383–407.
- Shikhare, S., Heim, S., Klein, E., Huber, S., & Willmes, K. (2015). Processing of numerical and proportional quantifiers. *Cognitive Science*, 39(7), 1504–1536.
- Slim, M. S., Lauwers, P., & Hartsuiker, R. J. (2021). Monolingual and bilingual logical representations of quantificational scope: Evidence from priming in language comprehension. *Journal of Memory and Language*, 116, 104184. <https://doi.org/10.1016/j.jml.2020.104184>
- Slobin, D. (2001). Form-Function Relations: How do Children Find out What they Mean?. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 406–449). Cambridge University Press.
- Slobin, D. I. (1985). Crosslinguistic Evidence for the Language-Making Capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 1157–1256). Lawrence Erlbaum Associates.
- Smoczyńska, M. (2008). *The test for the reception of grammar (Polish translation)*. Unpublished manuscript. Jagellonian University.
- Squire, C., Andrews, M., Davis, M., Esin, C., Harrison, B., Hyden, L. C., & Hyden, M. (2014). *What is narrative research?*. Bloomsbury Academic.
- Szustrowa, T., & Jaworowska, A. (2003). *Test matryc Ravena – wersja kolorowa*. Pracownia Testów Psychologicznych PTP.
- Thordardottir, E. (2015). Proposed Diagnostic Procedures for Use in Bilingual and Cross-Linguistic Contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 331–356). Multilingual Matters.
- Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages?. *Journal of Multilingual Communication Disorders*, 4(1), 1–21.
- Tomasello, M. (1992). *First verbs: A case study of early lexical development*. Cambridge University Press.
- Von Fintel, K., & Matthewson, L. (2008). Universals in semantics. *The Linguistic Review*, 25(1–2), 139–201.
- Wang, D., Lam, J. H. Y., McMillen, S., Su, P. L., Iglesias, A., Bedore, L. M., & Peña, E. D. (2025). Dual language profiles in Spanish–English bilingual children with and without developmental language disorder. *Bilingualism: Language and Cognition*, 29(459–474). <https://doi.org/10.1017/S1366728925100102>
- Zurer-Pearson, B. (2010). We can no longer afford a monolingual norm. *Applied Psycholinguistics*, 31(2), 339–343. <http://doi.org/10.1017/S014271640999052>