BUCLD 47 Proceedings
To be published in 2023 by Cascadilla Press
Rights forms signed by all authors

The Validity of a Transcript-Based Measure of Child Language Development in Czech

Klára Matiasovitsová, Petra Čechová, Jakub Sláma, Jolana Treichelová, and Filip Smolík

1. Introduction

The Index of Productive Syntax (IPSyn; Scarborough 1990) is a well-established measure of language development, widely used for English and adapted for other languages (Hewitt et al. 2005, for English; Nieminen 2009, for Finish, Oetting et al. 2010, for African-American English; Ooi and Wong 2012, for bilingual Chinese-English speaking children; Saban-Dülger et al. 2022, for Turkish, Washington et al. 2019, for English and Jamaican Creole). IPSyn examines whether specific morphosyntactic phenomena are productive in children's talk, regardless of their frequency of occurrence or the appropriateness of their use. The original IPSyn scale (Scarborough 1990) contains 56 categories divided in four subscales (*Noun Phrases*, *Verb Phrases*, *Questions and Negations*, *Sentence Structures*), and the child can score 0–2 points in each of them (0 = no occurrence; 1 = 1 occurrence; 2 = 2 or more distinct occurrences). The upper limit of two occurrences was set to simplify the calculation, and it presumably reflects the child's ability to use specific linguistic structures (Scarborough 1990).

Research has shown that there is a relation between IPSyn and age (Ooi and Wong 2012; Saban-Dülger et al. 2022; Scarborough 1990), although in children older than 4 years the results are more ambiguous (Oetting et al. 2010; Washington et al. 2019). That is consistent with the fact that IPSyn was designed for children between 24 and 48 months, and it evaluates the emergence of certain morphosyntactic phenomena in language production, rather than their mastery (Oetting et al. 2010). IPSyn also shows good concurrent validity with other transcript-based measures, namely Mean Length of Utterance (MLU; Saban-Dülger et al. 2022; Scarborough 1990) and Developmental Sentence Scoring (DSS; Saban-Dülger et al. 2022), as well as language tests (see Saban-Dülger et al. 2022; however, research by Condouris et al. 2003 did not confirm this).

* Klára Matiasovitsová, Charles University, Faculty of Arts, klara.matiasovitsova@ff.cuni.cz, Petra Čechová, Charles University, Faculty of Arts, petra.cechova@ff.cuni.cz, Jakub Sláma, Charles University, Faculty of Arts, slama@ujc.cas.cz, Jolana Treichelová, Charles University, Faculty of Arts, jolana.treichelova@ff.cuni.cz, Filip Smolík, Charles University, Faculty of Art, smolik@praha.psu.cas.cz. This study was supported by the project Grant Schemes at CU (reg. no. CZ.02.2.69/0.0/0.0/19_073/0016935).

Research has also revealed that IPSyn can be used to distinguish between children with typical and atypical language development (Condouris et al. 2003, for children with autism; Hewitt et al. 2005, for monolingual children with SLI; Ooi and Wong 2012, for bilingual children with SLI; Yang et al. 2022, for children with language delay). It is believed to be useful for intervention planning, as it examines the use of specific linguistic structures (Yang et al. 2022). Automated tools for IPSyn scoring simplify and expand the possibilities of its use by both clinicians and researchers (Finestack et al. 2020; Hassanali et al. 2014).

However, not all subscales of IPSyn seem to be equally useful. The subscale *Questions and Negations* has been identified as the most problematic, because it was not sensitive to age differences (Saban-Dülger et al. 2022; Washington et al. 2019; Yang et al. 2022) and did not differentiate between typically and atypically developing children (Yang et al. 2022). This subscale was even excluded in the study by Hewitt et al. (2005) because there were too few examples relevant to this scale in the transcripts due to the conversation settings (retelling stories and answering related questions).

Other subscales also have limitations. Scores on the *Noun Phrases* and the *Verb Phrases* subscales did not increase after 36 months and 42 months, respectively, in the study by Scarborough (1990). These results were confirmed in the studies by Hewitt et al. (2005), with ceiling effects observed for both subscales in 6-year-old children, and by Yang et al. (2022) for the *Noun Phrases* subscale. On the other hand, Yang et al. (2022) found the subscales *Verb Phrases* and *Sentence Structures* to be stable and differentiate between typical and atypical children. The *Sentence Structures* subscale differentiates between typically and atypically developed children even of school age (Hewitt et al. 2005).

This paper describes the adaptation of IPSyn for Czech, a highly inflected Slavic language. To our knowledge, it is only the second Slavic language for which IPSyn has been adapted (besides Russian, see Chernobilsky 2009). We focus on the following questions:

1. What is the reliability of the Czech adaptation of IPSyn?
2. How does the Czech adaptation of IPSyn relate to other measures of language development in Czech?
3. What are the psychometric properties of the subscales and specific structures in the Czech adaptation of IPSyn?

2. Method

2.1. Participants and procedures

For calculating IPSyn we used transcripts of recordings of 110 children, which are part of the unpublished corpus Labels2018, which contains transcripts collected in two longitudinal experimental studies, whose results were reported by Chromá and Smolík (2017) and Smolík and Bláhová (2021). Children in both of these studies were recorded for about 20 minutes during free play with their parents or other adults in a laboratory at the Institute of Psychology of the Czech

Academy of Sciences in Prague. All 110 children were recorded on average at 2;5,28 years (sd = 45 days) for the first time. The second round of data was collected at slightly different ages in the two studies: for 60 children, the second recordings were recorded at the average age of 3;7,17 (sd = 38 days), on average. For the remaining 50, the mean age at the time of the second round of data collection was 4;3,25 years (sd = 34 days). In total, the mean age of the 110 children at the second time point was 3;11,09 (sd = 130 days). At the first data point, the recordings included 24,658 words produced by children. At the second data point, children produced 33,033 words.

On both occasions, the participants were also given vocabulary and grammar comprehension tasks. However, three children did not participate in the grammar test at the first time point, and one did not participate at the second time point. The vocabulary test was not completed by three children at the first time point and by one from each (younger and older) group at the second time point.

The transcripts of the recordings were automatically lemmatized and tagged with MorphoDiTa, a freely available tool used for the morphological annotation of Czech corpora (Straková et al. 2014). Using custom-made Python scripts, the output of this tool was converted to the CHAT format used in the CHILDES database, to be analyzed by CLAN (MacWhinney 2000).

2.2. Measures

2.2.1. Transcript-based measures

We adapted IPSyn for Czech on the basis of its revised version by Altenberg et al. (2018) and Chernobilsky's (2009) modification of the original version for Russian. Our adaptation involves 53 items (instead of 59 in the version of Altenberg et al. 2018) divided into four subscales, as in the original English version: *Noun Phrases* (NP; with 20 items instead of 11 in Altenberg et al. 2018), *Verb Phrases* (VP; with 18 items instead of 17), *Questions and Negations* (QN; with 7 items instead of 11), and *Sentence Structures* (SS; with 8 items instead of 20).

Compared to the original list, we have removed items concerning structures that do not exist or are not typical in Czech, such as articles, the progressive suffix, or word order inversion in questions. We have also substituted the categories "any other bound morpheme on noun or adjective" and "any other bound morpheme on verb" with more specific items relating to verbal and nominal inflectional morphology that are not incorporated in the original version, in part because the respective phenomena often do not exist in English. The added items included the comparative and the superlative of adjectives and adverbs (N11 and N12 in Table 1), the different case forms of nouns (N13–15) and adjectives (N16), the reflexive pronoun (N20), verb conjugations in the indicative mood (V16–18), and verbs in the conditional and the imperative mood (V13–14). The *Sentence Structures* subscale was shortened, excluding categories that would require manual annotation of the entire corpus, such as the presence of a relative clause, an infinitive clause, or an *if*-clause.

A concise overview of the items included in the final version is provided in Table 1. For a more detailed explanation of the items see an extended version of the table available at https://doi.org/10.17605/OSF.IO/E795Q.

Table 1. Adaptation of IPSyn for Czech. ON = overall number; SN = subscale number.

ON	SN	Description	Example
Noun Phrases			
1	N1	noun (common and proper)	*pes* 'dog', *Pavel* 'Paul'
2	N2	prounoun (excluding modifiers)	*já* 'I', *co* 'what'
3	N3	modifier (adjectives, possessives, demonstratives, quantifiers)	*velký* 'big', *tvoje* 'your', *druhý* 'second', *tenhle* 'this'
4	N4	modifier + noun	*malá kočka* 'small cat'
5	N5	demonstrative + noun	*to kolo* 'that bike'
6	N6	preposition + modifier + noun	*v mém pokoji* 'in my room'
7	N7	plural noun (including pluralia tantum)	*hračky* 'toys', *nůžky* 'scissors'
8	N8	(modifier + noun) + verb / verb + (modifier + noun)	*tahle holčička má* 'this girl has', *vidí velký dům* 'sees a big house'
9	N9	modifier + modifier + noun	*ta bílá hračka* 'that white toy'
10	N10	adverb + modifier	*hodně malý* 'very small'
11	N11	comparative adjective/adverb	*milejší* 'nicer', *rychleji* 'faster'
12	N12	superlative adjective/adverb	*nejmilejší* 'nicest', *nejrychleji* 'fastest'
13	N13	noun in the nominative/accusative case	*židle* 'chair', *hračku* 'toy'
14	N14	noun in the locative/instrumental case	*pokoji* 'room', *lampou* 'lamp'
15	N15	noun in the genitive/dative case	*ovečky* 'sheep', *drakovi* 'dragon'
16	N16	adjective in one of these cases: genitive/dative/locative/instrumental	*roztomilého* 'cute', *velkému* 'big', *šťastném* 'happy', *malým* 'small'
17	N17	possessive adjective	*tátův* 'dad's'
18	N18	multiplicative numeral	*pětkrát* 'five times'

19	N19	quantifier + noun/pronoun in the genitive case	*pět psů* ‘five dogs’
20	N20	reflexive pronoun	*se, si* ‘-self, -selves’
Verb Phrases			
21	V1	verbs (lexical and modal)	*hraje si* ‘plays’, *chci* ‘want’
22	V2	preposition	*v* ‘in’, *k* ‘to’
23	V3	prepositional phrase	*ve škole* ‘at school’
24	V4	copula linking two nominals / a nominal and a predicative adjective	*tohle není stůl* ‘this is not a table’ *Ema je chytrá* ‘Ema is smart’
25	V5	adverb	*dobře* ‘well’
26	V6	modal verb + infinitive	*musí spinkat* ‘must sleep’
27	V7	3SG verb in the present tense	*pije* ‘drinks, is drinking’
28	V8	past modal verb + infinitive	*mohli číst* ‘could read’
29	V9	past participle (lexical or modal)	*dělala* ‘did’, *mohli* ‘could’
30	V10	past tense auxiliary *být* ‘to be’	*jsem myslel* ‘I thought’
31	V11	copular or lexical *být* ‘to be’ in the past tense	*jsi byl* ‘you were’
32	V12	future tense auxiliary *být* ‘to be’	*budeš spát* ‘will sleep’
33	V13	verb in the conditional mood	*zpíval by ‘would sing’*
34	V14	verb in the imperative mood	*řekni* ‘say’
35	V15	lexical verb + infinitive	*potřebuju pít* ‘I need to drink’
36	V16	1SG verb form (lexical or modal)	*piju* ‘I drink’
37	V17	2SG verb form (lexical or modal)	*piješ* ‘you drink’
38	V18	plural verb form (lexical or modal)	*pijí* ‘they drink’
Questions and Negations			
39	Q1	intonationally marked question	*spí?* ‘she sleeps?’
40	Q2	*wh*-word question	*proč?* ‘why?’
41	Q3	negation particle / negative pronoun or pronominal adverb	*ne* ‘no’, *nic* ‘nothing’, *nikde* ‘nowhere’

42	Q4	*wh*-word question containing a verb	*proč se směje?* ‘why is she laughing?’
43	Q5	nominal phrase + negated lexical verb	*já nevím* ‘I don’t know’
44	Q6	negated auxiliary/modal verb	*nebudu jíst* ‘I won’t eat’, *nemůžu* ‘I can’t’
45	Q7	pronominal adverbs corresponding to *why*/*when*/*which*/*whose*	*proč*, *kdy*, *který*, *čí*
Sentence Structures			
46	S1	two-word combination	*já vím* ‘I know’
47	S2	subject + verb	*táta zpívá* ‘daddy is singing’
48	S3	verb + object	*jím zmrzlinu* ‘I’m eating ice-cream’
49	S4	subject + verb + object	*máma vaří oběd* ‘mommy is cooking lunch’
50	S5	conjunction corresponding to *and*/*or*	*a* ‘and’, *nebo* ‘or’, *ani* ‘nor’, *i* ‘and’, *či* ‘or’
51	S6	coordinating conjunction connecting clauses	*jím a koukám na televizi* ‘I’m eating and watching the TV’
52	S7	subordinating conjunction connecting clauses	*nevím, co říkala* ‘I don’t know what she said’
53	S8	sentence with three or more verbal phrases	*smála se, protože zpíval a tancoval* ‘she laughed was laughing because he was singing and dancing’

IPSyn was calculated by the CLAN tool with the following command: *ipsyn +lcze +c[number of utterances] file.cha*. The script used to calculate IPSyn (see also at https://doi.org/10.17605/OSF.IO/E795Q) was adapted from the English version and, like the original, ensured that a given category received the maximum score of two points only if the two examples were sufficiently distinct from each other (e.g., plural inflection was used with two different nouns). For the calculation of IPSyn, we used all complete and intelligible utterances produced by the child, similarly to Oetting et al. (2010) and Hewitt et al. (2005); the latter study, nevertheless, excluded children with less than 50 utterances.

In addition, we calculated MLU and the number of different words (NDW) for all the transcripts using the CLAN utility. MLU of all utterances was measured in words. Incomplete, unintelligible and immediately repeated utterances, as well as false starts, repetitions, or nonlinguistic sounds were not included in the calculation. The number of different words was calculated as the number of

different lemmas, i.e., different morphological forms of the same lexeme were counted as one word.

2.2.2. Test-based measures

To establish the criterion validity of IPSyn, we used as additional measures scores from tests of vocabulary and grammar. The vocabulary task was a picture comprehension test with a format similar to the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn 2007), and the grammar comprehension task was similar to the Test for Reception of Grammar (TROG; Bishop 2003). At the first time point, the tests were the same for all the children, and so was the grammar test at the second time point. The vocabulary task at the second time point differed for the younger and the older group.

2.3. Psychometric analyses

Validity and reliability estimates, as well as the item analysis, were calculated using the online application ShinyItemAnalysis (http://www.shinyitemanalysis.org/; Martinková and Drabinová 2018).

4. Results

4.1. Reliability and validity of IPSyn

The internal consistency of the total IPSyn score was estimated using Cronbach's alpha, yielding the values 0.94 and 0.89 for the first and second time point, respectively, which indicates very good internal consistency of the measure.

We also found strong correlations between IPSyn and MLU, and between IPSyn and NDW at both time points, with stronger results for the first time point (r's > 0.87; see Table 2). However, the most remarkable results are the moderate correlations between IPSyn and both of the test-based measures, also stronger at the first time point (r's > 0.34). At the second time point, the correlations between IPSyn and other measures were always lower in the children tested at 4;3 years than those tested at 3;7 years.

Table 2. Correlations between IPSyn and other transcript-based and test-based measures at both time points. GRAM = grammar; VOC = vocabulary. The two numbers separated by a slash refer to the two subgroups: the first to the 60 children tested at 3;7 years, and the second to the 50 children tested at 4;3 years. The critical value of the correlation coefficient *r* for α = 0.05 is 0.19 for 110 children, 0.26 for 60 children and 0.28 for 50 children.

	MLU	NDW	GRAM	VOC
IPSyn 1	0.88	0.90	0.35	0.38
IPSyn 2	0.77 (0.81/0.74)	0.82 (0.87/0.71)	0.35 (0.38/0.31)	0.33/0.28

The predictive validity of IPSyn was examined using linear regression models that predicted IPSyn at the second time point from IPSyn, from the comprehension tasks, or from all measures at the first time point. The results of the three models are shown in Table 3. In a model that included all three measures, IPSyn scores at 2;5 years predicted IPSyn scores at 3;11 with $\beta = 0.37$, above and beyond the effects of the vocabulary and grammar comprehension tasks. Vocabulary also showed a significant unique effect ($\beta = 0.26$).

Table 3. Regression models showing the unique effects of IPSyn, grammar, and vocabulary measured at the first time point on IPSyn results at the second time point. GRAM = grammar; VOC = vocabulary.

	IPSyn 2					
Predictors	*std. Beta*	*p*	*std. Beta*	*p*	*std. Beta*	*p*
(Intercept)	0.00	**<0.001**	-0.00	**<0.001**	-0.00	**<0.001**
IPSyn 1	0.50	**<0.001**			0.37	**<0.001**
GRAM 1			0.21	**0.020**	0.11	0.207
VOC 1			0.38	**<0.001**	0.26	**0.004**
Observations	110		106		106	
R^2 / R^2 adjusted	0.252 / 0.245		0.224 / 0.209		0.329 / 0.309	

As IPSyn is considered primarily a measure of grammar, we also used regression models to examine the unique effects of IPSyn and the results of the test-based tasks measured at the first time point on the results of the grammar task

at the second time point. As shown in Table 4, IPSyn at 2;5 years predicted results in the grammar task at 3;11 years above and beyond the effects of other predictors, and with the greatest effect size ($\beta = 0.36$ as opposed to $\beta = 0.28$ and 0.23 for the grammar task and the vocabulary task, respectively). This indicates a specific relation between grammatical development in production and later comprehension of grammar.

Table 4. Regression models showing the unique effects of IPSyn, vocabulary, and grammar measured at the first time point on grammar results at the second time point. GRAM = grammar; VOC = vocabulary.

GRAM 2			
Predictors	*std. Beta*	*Statistic*	*p*
(Intercept)	0.00	0.48	0.633
IPSyn 1	0.36	4.20	**<0.001**
GRAM 1	0.28	3.45	**0.001**
VOC 1	0.23	2.77	**0.007**
Observations		105	
R^2 / R^2 adjusted		0.410 / 0.392	

4.2. Reliability and validity of the subscales of IPSyn

To examine the structure of IPSyn and the contribution of the four subscales, we calculated reliability and validity indices for these subscales. The results are summarized in Table 5. Cronbach's alpha for IPSyn decreased when the NP, the VP or the SS subscale was excluded, but it was minimally affected by the exclusion of the QN subscale, indicating that this subscale does not contribute reliably to the total score.

The validity of the individual subscales was estimated by examining correlations between the subscale scores on the one hand, and the total IPSyn score, grammar comprehension, or vocabulary comprehension scores on the other. The subscale-test correlations are generally high, with the lowest values for the QN subscale (see Tables 5 and 6). The relations between test scores and IPSyn are moderate, and again the QN subscale shows the lowest values. The IPSyn-test correlations are generally higher at the first time point, indicating that IPSyn may be a more sensitive measure at around the age of 2.5 than at 4 years of age.

Table 5. Reliability and (criterion) validity (with grammar and vocabulary as criterion variables) of the IPSyn subscales at the first time point. *α*-drop = Cronbach's alpha of the test without a given subscale; *r* subscale-test/grammar/vocabulary = Pearson correlation coefficients.

subscale	*α-drop*	*r subscale-test*	*r subscale-grammar*	*r subscale-vocabulary*
NP	0.81 (GRAM) 0.80 (VOC)	0.92	0.31	0.42
VP	0.81	0.96	0.37	0.33
QN	0.88	0.80	0.21	0.26
SS	0.85	0.89	0.33	0.34
Cronbach's alpha for the test as a whole: 0.88.				

Table 6. Reliability and (criterion) validity (with grammar and vocabulary as criterion variables) of IPSyn subscales at the second time point. The two numbers divided separated by a slash refer to the two subgroups, the first to the 60 children tested at 3;7 years, and the second to the 50 children tested at 4;3 years. *α*-drop = Cronbach's alpha of test without given subscale; *r* subscale-test/grammar/vocabulary = Pearson correlation coefficients.

subscale	*α-drop*	*r subscale-test*	*r subscale-grammar*	*r subscale-vocabulary*
NP	0.72	0.82	0.38	0.27/0.36
VP	0.69	0.89	0.29	0.31/0.17
QN	0.78	0.69	0.13	0.34/0.04
SS	0.74	0.77	0.29	0.20/0.11
Cronbach's alpha for the test as a whole: 0.79.				

4.3. Item analysis of IPSyn

Figures 1 and 2 show the difficulty and discrimination values for individual IPSyn items: difficulty is the proportion of scoring in the item, with 1 equaling the perfect performance and thus indicating an easy item, and 0 equaling no

performance thus indicating a difficult item. The discrimination value is the correlation between the item score and the total IPSyn score. Items with high or low difficulty have decreased discrimination values because their variance is limited, but these items may be useful for children who are at a different overall level of development. However, if an item is of moderate difficulty and yet of low or moderate discrimination, this indicates that the item does not reflect the same psychometric property as the rest of the scale.

The analysis revealed low discrimination in a few items at one or both time points. Items Q3 (negation particle / negative pronoun or pronominal adverb) and Q6 (negated auxiliary/modal verb) were problematic at each time point. At the second time point, items N7 (plural noun), N15 (noun in the genitive/dative case) and V10 (past tense auxiliary *být* 'to be') were problematic, too. The low discrimination of the two items from the *Questions and Negations* subscale is in line with the limited reliability and criterion validity identified in the previous analyses. A possible explanation might be that the discourse during play did not provide enough opportunities for the use of negation (which is probably a shortcoming of the whole subscale of *Questions and Negations*). This may also be true for the past tense auxiliary *být* 'to be' (V10), as free play in the laboratory probably provides few contexts for the use of the past tense. The low informativeness of the item N7 (plural noun) was explained by Yang et al. (2022) by its early acquisition.

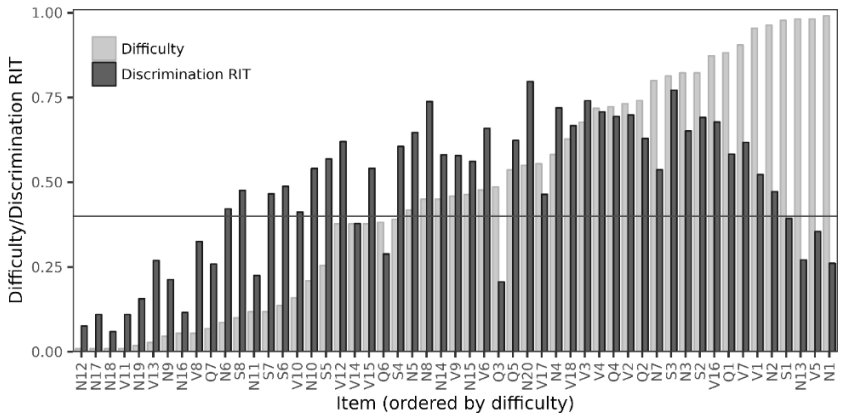

Figure 1. Item analysis of IPSyn at the first time point. Difficulty = item difficulty estimated as an average item score divided by its range; discritimination RIT = Pearson correlation between item and total score.

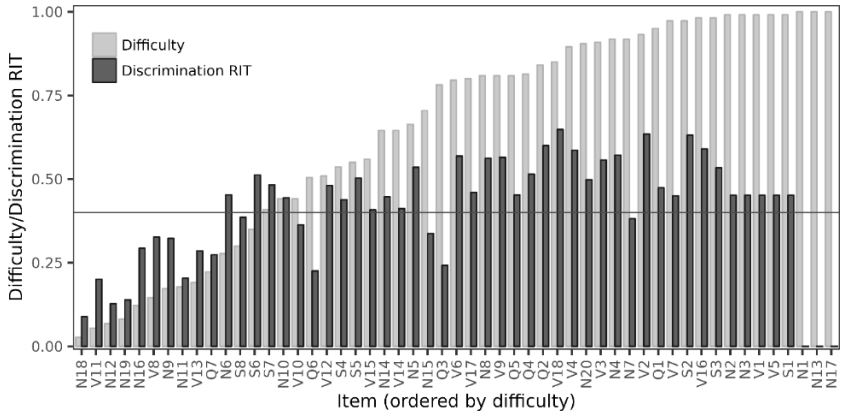

Figure 2. Item analysis of IPSyn at the second time point. Difficulty = item difficulty estimated as an average item score divided by its range; discritimination RIT = Pearson correlation between item and total score.

5. Discussion

The Czech adaptation of IPSyn demonstrated good internal consistency at both time points, indicating adequate reliability of the measure. The validity of IPSyn was supported by concurrent relations not only to other transcript-based measures (also reported by Nieminen 2009, for MLU and Washington et al. 2019, for MLU and NDW), but also to test-based measures focused on receptive skills in children (which is consistent with the results of Saban-Dülger et al. 2022, but contrary to the findings of Condouris et al. 2003, who only found significant correlations between IPSyn and transcript-based, not test-based measures in children with autism between 4 and 14 years of age). Lower correlations between IPSyn and other measures at the second time point (especially in children tested at 4;3 years) together with the lower subscale-test correlations may indicate lower validity of IPSyn in children at about 4 years old (see Oetting et al. 2010; Washington et al. 2019). IPSyn was predicted by the results in the vocabulary test and at the same time predicted the results in the grammar test. These results demonstrate relation between transcript-based and test-based measures.

The item analysis revealed the low informativeness of some items, especially of two of the seven structures in the subscale *Questions and Negations*. The correlations with test measures were also lower for this subscale at both time points, and its exclusion from the total IPSyn score did not result in a decrease in reliability. This supports the findings reported by Washington et al. (2019) and Yang et al. (2022), who attributed this result to pragmatic effects, especially the role of context in conversation: the presence or absence of questions and negators in a conversation may depend to a large extent on the situation, and even advanced children may have limited opportunities to use them. According to Saban-Dülger

et al. (2022), these results may also be due to the lower number of items in the *Questions and Negations* subscale.

In contrast to the findings by Scarborough (1990) and Yang et al. (2022), the *Noun Phrases* subscale was not less informative than that of *Verb Phrases* and *Sentence Structures*. On the contrary, its good reliability and validity were confirmed by its highest Cronbach's alpha, its subscale-test correlations, and its good criterion validity at both time points. This might be explained by the more complex morphosyntactic system of Czech nominal phrases when compared to other languages such as English. Hence, the acquisition of structures of the *Noun Phrases* subscale might be more difficult in Czech, reducing the possibility of a ceiling effect. Another possible explanation is the number of structures in this subcategory – the Czech adaptation of IPSyn included 20 items in the *Noun Phrases* subscale instead of 11 in the version of IPSyn by Altenberg et al. (2018). The higher reliability of subscales with a higher number of items was pointed out by Scarborough (1990).

To summarize, we proposed and tested a system for the calculation of IPSyn in a morphologically complex language that has not been studied previously with respect to this measure. The results confirm that the system provides reliable and valid indicators of language development, and has good potential for use in research and intervention concerning young Czech children.

References

Altenberg, Evelyn P., Jenny A. Roberts & Hollis S. Scarborough. 2018. Young children's structure production: A revision of the index of productive syntax. *Language, Speech, and Hearing Services in Schools* 49(4): 995-1008.

Bishop, Dorothy V. M. 2003. *Test for reception of grammar, version 2 (TROG 2)*. London, UK: Pearson Assessment.

Condouris, Karen, Echo Meyer & Helen Tager-Flusberg. 2003. The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech and Language Pathology* 12(3): 349-358.

Dunn, Lloyd M., and Douglas M. Dunn. 2007. *Peabody Picture Vocabulary Test–Fourth Edition*. American Guidance Service.

Finestack, Lizbeth H., et al. 2020. Using computerized language analysis to evaluate grammatical skills. *Language, Speech, and Hearing Services in Schools* 51(2): 184-204.

Hassanali, Khairun-nisa, et al. 2014. Automatic generation of the index of productive syntax for child language transcripts. *Behavior research methods* 46(1): 254-262.

Hewitt, Lynne E., et al. 2005. Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of communication disorders* 38(3): 197-213.

Chernobilsky, Ellina D. 2009. *Linguistic competence of five and six year olds: Analysis of narrative samples of Russian, English and Russian-English bilingual speakers*. Rutgers The State University of New Jersey-New Brunswick.

Chromá, Anna & Filip Smolík. 2017. Personal pronouns and verb person inflections: Relations with linguistic development and social understanding. In LaMendola M.,

Scott J. (eds.), *Proceedings of the 41st annual Boston University Conference on Language Development*, 168-178. Somerville, MA: Cascadilla Press.
MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
Martinková, Patrícia & Adéla Drabinová. 2018. ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal* 10(2): 503-515.
Nieminen, Lea. 2009. MLU and IPSyn measuring absolute complexity. *Eesti Rakenduslingvistika Ühingu aastaraamat* 5: 173-185.
Oetting, Janna B., et al. 2010. Index of productive syntax for children who speak African American English. *Language, Speech, and Hearing Services in Schools* 41: 328-339.
Ooi, Carmen C-W. & Anita M-Y Wong. 2012. Assessing bilingual Chinese–English young children in Malaysia using language sample measures. *International Journal of Speech-Language Pathology* 14(6): 499-508.
Saban-Dülger, Nur Seda, Figen Turan & Esra Özcebe. 2022. The Adaptation of Developmental Sentence Scoring and Index of Productive Syntax to Turkish. *Journal of Speech, Language, and Hearing Research* 65(3): 1001-1024.
Scarborough, Hollis S. 1990. Index of productive syntax. *Applied psycholinguistics* 11(1): 1-22.
Smolík, Filip & Veronika Bláhová. 2021. Early linguistic reference to first and second person depends on social understanding as well as language skills: Evidence from Czech 30-month-olds. *First Language* 41(1): 109-125.
Straková, Jana, Milan Straka & Jan Hajič. 2014. Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13-18. Baltimore, Maryland: Association for Computational Linguistics.
Washington, Karla N., et al. 2019. Bilingual preschoolers' spontaneous productions: Considering Jamaican Creole and English. *Language, Speech, and Hearing Services in Schools* 50(2): 179-195.
Yang, Ji Seung, Brian MacWhinney & Nan Bernstein Ratner. 2022. The Index of Productive Syntax: Psychometric properties and suggested modifications. *American journal of speech-language pathology* 31(1): 239-256.