

Vodickova, Katerina; Kostelecka, Yvona

Diagnostic tests in Czech for pupils with a first language different from the language of schooling

CEPS Journal 6 (2016) 1, S. 31-48



Quellenangabe/ Reference:

Vodickova, Katerina; Kostelecka, Yvona: Diagnostic tests in Czech for pupils with a first language different from the language of schooling - In: CEPS Journal 6 (2016) 1, S. 31-48 - URN: urn:nbn:de:0111-pedocs-120475 - DOI: 10.25656/01:12047

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-120475>

<https://doi.org/10.25656/01:12047>

in Kooperation mit / in cooperation with:



University of Ljubljana
Faculty of Education

<http://www.pef.uni-lj.si>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/3.0/de/deed> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/3.0/de/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Diagnostic Tests in Czech for Pupils with a First Language Different from the Language of Schooling

KATEŘINA VODIČKOVÁ*¹ AND YVONA KOSTELECKÁ²

☞ Mastering a second language, in this case Czech, is crucial for pupils whose first language differs from the language of schooling, so that they can engage more successfully in the educational process. In order to adjust language teaching to pupils' needs, it is necessary to identify which language skills or individual competences set out within the framework of communicative competence should be developed. For this purpose, a new diagnostic test for lower and upper graders of primary schools was designed. Although it is not a high-stakes test, it is essential that its validity, reliability and practicality are ensured, as well as its positive impact on the teaching process, pupils, teachers, schools and society. The present paper introduces the position of pupils with a first language other than Czech in the Czech Republic. It presents a recently developed diagnostic tool and documents the characteristics of the test, such as validity, reliability, impact and practicality.

Keywords: Common European Framework of Reference, Czech as a second language, diagnostic test, young learners

1 *Corresponding Author. Charles University in Prague, Institute for Language and Preparatory Studies, Czech Republic; katerina.vodickova@ujop.cuni.cz

2 Charles University in Prague, Pedagogical Faculty, Czech Republic

Diagnostični testi na Češkem za učence, katerih prvi jezik ni enak jeziku šolanja

KATEŘINA VODIČKOVÁ* IN YVONA KOSTELECKÁ

Obvladanje drugega jezika, v tem primeru češčine, je ključnega pomena za učence, katerih prvi jezik ni enak jeziku šolanja, saj se le tako lahko uspešno vključijo v vzgojno-izobraževalni proces. Da lahko prilagodimo poučevanje jezika potrebam učencev, je nujno prepoznati, katere jezikovne spretnosti ali individualne zmožnosti, določene v okviru komunikacijskih zmožnosti, morajo biti razvite. V ta namen je bil oblikovan nov diagnostični test za učence nižjih in višjih razredov osnovne šole. Kljub splošnosti testa je treba zagotoviti, da ima ustrezno veljavnost, zanesljivost in praktičnost ter da ima pozitiven vpliv na proces poučevanja, učence, učitelje, šole in na družbo. V prispevku je predstavljen položaj učencev na Češkem, katerih prvi jezik ni češčina. Predstavljeni so pred kratkim razvito diagnostično orodje in karakteristike testa, kot so: veljavnost, zanesljivost, vpliv in praktičnost.

Ključne besede: skupni evropski referenčni okvir, češčina kot drugi jezik, diagnostični test, mlajši učenci

Introduction

As a consequence of the integration of the Czech Republic into the European Union, and of continuing globalisation, we have witnessed an increase in the migration of population in the last decades. Until 1989, Czechoslovakia was characterised by emigration, but after the Velvet Revolution this country of emigration began turning into a country of immigration (Drbohlav, 2011). As a result of this change it became necessary for several institutions, and in particular integration policy, to adapt the prevailing attitude in a relatively short time.

The process of integration into the host society is influenced by a range of factors, many of which are already the subject of detailed research, such as the institutional environment and migration policy (Heckmann & Schnapper, 2003), confession (Foner & Alba, 2008), cognitive skills (Suárez-Orozco, 2007) and mastering the language³ (Chiswick & Miller, 2001). Some of these factors have already been examined in the Czech context, as well (cf., e.g., Drbohlav, 2011; Janská et al., 2011).

In connection with the growing number of non-native speakers in the Czech Republic, there has been an increasing interest in the application and study of the Czech language, not only as a foreign language but also as a second language. The growing number of children of migrants⁴ (i.e., pupils with a first language, hereafter L1, that is different from the language of schooling) at Czech schools places greater demands on teachers, and therefore also necessitates a more systematic approach for pedagogical workers when solving basic linguo-didactic issues in multi-cultural classes at primary schools (cf., e.g., Šindelářová & Škodová, 2013). Although mastering a second language becomes a prerequisite for accessing and completing education, as well as for integration into the school group and consequently into society as a whole (cf. Kostelecká et al., 2013, p. 7), children of migrants face a rather complicated situation at Czech schools. The Czech school system lacks longstanding practical experience of teaching Czech as a second language, and of integrating children of migrants into the educational process and teaching multicultural classes.

We have already mentioned that mastering a second language has social and practical significance for children of migrants, and is therefore crucial for successful integration.

3 We understand the term mastering a language as an umbrella term for learning a language and language acquisition.

4 In the 2010–2011 academic year, children of migrants represented 1.4% of the total number of children in Czech preschool facilities, while also constituting 1.7% of all elementary school pupils and 1.5% of grammar school pupils (according to the Statistical Yearbook of Education 2010/11).

In the Czech Republic, the level of communicative competence and language skills with which children of migrants arrive has not yet been measured. In order to acquire this and other information, a diagnostic test for first- and second-grade pupils at primary schools has been developed at the Pedagogical Faculty, Charles University in Prague. The present paper aims to discuss this diagnostic instrument, including its basic characteristics and intended impact.

In Part 2, we briefly address testing in general, with an emphasis on diagnostic tests and the specifics of testing young learners. We also explore the situation related to language testing, in particular testing young learners and diagnostic tests in the Czech Republic. The heart of the paper is constituted by Part 3, in which the developed diagnostic instrument is described, and Part 4, where we attempt to substantiate that it is a valid, reliable and practical instrument with a positive impact as a diagnostic tool. An outline of the direction in which the work with diagnostics might continue in the future is given in Part 5.

Developing diagnostic language tests

It is obvious that the assessment of language skills and competencies represents a very important component of language teaching. During the past decades, the field of assessment has developed considerably both theoretically and methodologically. Since language testing has become an integral part of teaching foreign languages and has developed into an individual branch of applied linguistics, there has been an increase in the quantity of publications and journals on language assessment (e.g., *Understanding Language Testing* by Dan Douglas), and in the number of specialised organisations such as the Association of Language Testers in Europe (ALTE), the European Association of Language Testing and Assessment (EALTA) and the International Language Testing Association (ILTA), as well as the creation of the Association of Language Testers AJAT in the Czech Republic in 2012. There has also been an increase in the number of conferences, workshops and seminars on this topic.

Despite this fact, there is little theoretical background and research on diagnostic testing, although publications such as *Diagnosing Foreign Language Proficiency* (2005) by Alderson have contributed considerably to the field of diagnostic language testing. Such monographs nonetheless remain scarce. In addition, especially in the case of Czech as a second/foreign language, the number of diagnostic tests in second/foreign languages is, to our knowledge, limited. Alderson et al. (2015, p. 237) point out “the scarcity of true diagnostic assessment” and believe this may be connected with “a lack of a theory of what diagnosis in [second/foreign language] actually entails”.

The definition of what constitutes a diagnostic test is itself problematic. Following the comparison of a number of definitions of this type of test, Alderson (2005) arrives at a set of features that diagnostic tests should demonstrate. They include, among others, the ability “to identify strengths and weaknesses in a learner’s knowledge and use of language” (p. 11), but place an emphasis on weaknesses, so that correction can be ensured during the subsequent teaching. These tests are mostly low- or no-stakes, and should therefore provide detailed feedback and enable thorough analysis. According to Alderson (2005), diagnostic tests are based either on content that has been covered in instruction, or on some theory of language development. Alderson (2005) also points out that achievement tests and proficiency tests are often used for diagnostic purposes, or diagnostic tests are used for placement purposes.

Harding et al. (2015) developed a set of principles for diagnostic assessment, which emphasise: a) the role of the user of the test who is responsible for the diagnosis, as opposed to the test itself; b) the importance of detailed feedback for the test-taker; c) the necessity of including a number of views, such as self-assessment; d) the role of various stages in diagnostic assessment, such as listening/observing; and e) the fact that diagnostic assessment should lead to remediation or tailor-made support. However, some of these principles are often omitted in practice, which, to a certain extent, also seems to be apparent in Czech diagnostic tests for children of migrants. In this specific case, the original use of the diagnostic test, as well as the continuous work on test development, should be taken into account.

Diagnostic testing in the Czech Republic

In language testing, we encounter various types of tests, differing largely in purpose and therefore in the interpretation of results. In the Czech context, these include proficiency tests (e.g., Czech Language Certificate Exam⁵); placement tests (offered for those interested in courses by most language schools, and provided for those who are interested in taking online courses, such as at the Institute for Language and Preparatory Studies, Charles University in Prague, hereafter ILPS CU); progress tests (continuous assessment verifying that the pupils/students have mastered the target material of teaching and learning; these tests have traditionally been a part of foreign language teaching at Czech primary, secondary and language schools); and achievement tests (e.g., the end-of-course examination in Czech at ILPS CU).

5 <http://ujop.cuni.cz/cce>

As mentioned above, diagnostic tests do not enjoy a long tradition in the Czech context, or, more precisely, we are missing available literature on Czech diagnostic testing of a second/foreign language. The Diagnostic handbook *Diagnostika úrovně znalosti českého jazyka* (Diagnosing the Level of Czech) was written to help professionals from the Centre for Integration to get a basic idea of the level of their clients' communicative competence in Czech. In this case, the diagnostic test is designed as a proficiency test of language skills and is intended for adult non-native speakers.

Comprehensive information on other diagnostic tests (diagnostic not only in name) has, however, been so far absent in the Czech Republic.

Testing young learners in the Czech Republic

In recent decades, considerably more attention has been paid to testing young learners than to diagnostic tests (cf., e.g., Hughes, 2003; Ioannou-Georgiou & Pavlou, 2003; McKay, 2010). It is obvious that testing young learners in a second/foreign language differs from testing adult language users; among other things, their ages, their cognitive, emotional, social and physical growth, their attention span and their literacy skills require significantly different approaches. The importance of positive motivation must also be considered.

Although foreign language tests represent a common part of teaching at Czech primary and secondary schools, diagnostic tests of the Czech language and tests of the Czech language as a second language are not common. This fact, along with the need to diagnose the level of communicative competence achieved by children of migrants, has, among other things, led to the development of the diagnostic instrument described in the following section.

Diagnostic tests of Czech for children of migrants

A suite of diagnostic tests for children of migrants was developed in the course of 2010–2014. Using an existing placement, achievement or proficiency test was not considered appropriate, primarily for the following reasons: a) the purpose of the test may vary; b) there is a lack of Czech language tests designed exclusively for young learners and, to our best knowledge, none for children of immigrants; c) even if they existed, using syllabus-based achievement tests would not take into account the fact that the children may have learned Czech from various sources, or without reference to official teaching materials at all (there is no specific syllabus that has to be covered before the test, or that should be covered afterwards); and d) the proficiency test Czech Language Certificate

Examination for Young Learners (CCE–A1 for Young Learners and CCE–A2 for Young Learners⁶) is subject to a fee, and, moreover, is available only at A1 and A2 levels according to the *Common European Framework of Reference for Languages* (2001, hereafter CEFR), as well as being too time consuming.

For these reasons, a pilot version of a tailor-made diagnostic test for primary schools was introduced in 2010. It was decided that the test should be a proficiency test, as there is no syllabus to which the test can relate. For this reason, there is no grammar or vocabulary test, although some information on the level of grammatical, lexical and other competencies can be inferred from the productive-skills subtests. It should also be born in mind that the first versions of the test were meant to be used to map the language situation among children of migrants in the Czech Republic, and were applied at a number of selected schools that were interested in taking part in the project and that are attended by larger numbers of children of migrants.

The format of the diagnostic test

Within project no. 13-32373S of the Czech Science Foundation, two diagnostic tests were developed. The first of these is aimed at lower graders. Taking into account the development of language skills in the respondents' first language and their cognitive development, this test is designed for pupils attending the 3rd, 4th and 5th grades, which roughly corresponds to the ages from 8 to 11. It verifies the level of communicative competence within language skills at the A1 and A2 levels according to the CEFR. The second test is aimed at upper graders, i.e., the age group between 12 and 16, and verifies the level of language skills at the A1, A2 and B1 levels according to the CEFR.

When designing the test, the test developers could not base it directly on the CEFR and its descriptors, as these are defined for adult language users and do not take into account children's cognitive development and the communicative situations they enter. The tests are therefore founded on documents based on the CEFR, that is, language portfolios: the diagnostic test for lower graders is based on the Portfolio for Learners Up to the Age of 11 (Nováková et al., 2001), and the test for upper graders is based on the European Language Portfolio for Learners aged 11 to 15 (Perclová & Marešová, 2001), which means that the Can Do Statements for the particular age groups serve as the basis for the specific aims that are verified within each subtest.

General information about the diagnostic test

The learners, both lower and upper graders, first take the lower level test in reading, listening and writing. If they pass, that is, if they achieve at least 60% in each subtest at this level, they proceed to the higher level test.

The scores are reported per subtest per level, as the original test is meant to map the level of communicative competence of children of migrants attending Czech primary schools. Negotiations are currently being held as to whether the test could serve as the basis for a tool to measure the progress of these pupils in Czech and/or their level of communicative competence in Czech, in order to determine how many extra lessons of Czech per week are necessary.

The format of the diagnostic test for lower graders

The lower-grader diagnostic test at the A1 and A2 levels verifies all four language skills in four subtests: reading, listening, writing and speaking. The pupils can gain a maximum of 15 points in each subtest per level (see Table 1).

Table 1. *The format of the lower-grader diagnostic test*

Level	Subtest	No. of tasks/ Total no. of items	No. of points	Time
A1	Listening	3/15	5+5+5	10 minutes
	Reading	3/15	5+5+5	12 minutes
	Writing	2	6+9	10 minutes
	Speaking	1	15	3 minutes
A2	Listening	3/15	5+5+5	15 minutes
	Reading	3/15	5+5+5	18 minutes
	Writing	2	6+9	15 minutes
	Speaking	1	15	5 minutes

The format of the diagnostic test for upper graders

The upper-grader diagnostic test verifies the level of communicative competence in four language skills at the A1, A2 and B1 levels according to the CEFR. The format of the test corresponds to the format of the diagnostic test for lower graders (cf. Table 2), although the test techniques may vary, as does

the time allotted to each subtest. It should be noted that there is only one task in the subtest Writing at the A2 and B1 level, in order to eliminate the error rate caused by fatigue and reduced concentration.

Table 2. *The format of the upper-grader diagnostic test*

Level	Subtest	No. of tasks/ Total no. of items	No. of points	Time
A1	Listening	3/15	5+5+5	6 minutes
	Reading	3/15	5+5+5	10 minutes
	Writing	2	5+10	10 minutes
	Speaking	1	15	3 minutes
A2	Listening	3/15	5+5+5	9 minutes
	Reading	3/15	5+5+5	10 minutes
	Writing	1	15	10 minutes
	Speaking	1	15	4-5 minutes
B1	Listening	2/15	5+10	13 minutes
	Reading	3/15	5+5+5	15 minutes
	Writing	1	15	15 minutes
	Speaking	1	15	4-5 minutes

The piloting phase, using the first version of the test, took place throughout 2010. After revisions were made based on the results and experience of the pilot, pretesting took place under the same test conditions in 2013. In order to ensure that both the piloted and pretested population were the same as the intended test population, the piloting and pretesting were realised at a number of primary schools on a voluntary basis. Only children between the 3rd and 9th grades whose first language was other than Czech were invited to take the test, based on parental consent.

Validity, reliability, impact and practicality of the diagnostic test

Validity, reliability, impact and practicality are usually considered the most essential quality indicators.

Validity

Validity, as an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment, is a crucial concept in language testing (cf., e.g., Hughes, 2003; Messick, 1989). Messick (1989) distinguishes six aspects of validity: content, substantive, structural, external, generalizability and consequential. In his view, the content aspect of construct validity includes evidence of content relevance, representativeness and technical quality.

In the case of the diagnostic test, these three components are addressed mainly by defining and adhering to the construct through detailed test specifications linked to the European Language Portfolios and through following these specifications.

Content validation (cf., e.g., Alderson, Clapham, & Wall, 1995; Hughes, 2003) of the test took place above all by gathering the opinions of independent experts. Four experts were asked to review the test sets, two of these experts were experienced in language testing and two in teaching young learners, while one also had experience in designing textbooks for young learners of Czech. All four experts were experienced in teaching Czech to foreigners, but they came from various backgrounds (university teachers, teaching Slavonic versus non-Slavonic students, teaching young learners versus adult learners, etc.). Their reviews included comparisons of the test content with the test specifications. The analysis showed that the difficulty levels, i.e., A1–B1, had been maintained; however, one of the reviewers recommended meeting the construct of certain language skills – specifically, writing and reading – so that they match the descriptors for the given level referred to in the corresponding European Language Portfolios, and so that the acquired language material could be considered representative. In a few cases, the reviewers recommended adapting the communication situation so that it would correspond more accurately to situations that the given age groups enter.

Adjusting the test on the basis of the aforementioned comments resulted in an increase in content validity and, consequently, improved probability that the test more accurately measures that which it declares to measure (cf. Hughes, 2003, p. 27).

Criterion-related validity “relates to the degree to which results on the test agree with those provided by some independent and highly dependable assessment of the candidate’s ability” (Hughes, 2003, p. 27). Much like, for example, Alderson, Clapham and Wall (1995) and Davies et al. (1999), Hughes (2003)

distinguishes between two types of criterion-related validity (external validity, in Alderson, Clapham and Wall's terminology): concurrent validity and predictive validity.

Concurrent validity is established by "the relationship between what is measured by a test ... and another existing criterion measure, which may be a well-established standardised test" (Davies et al., 1999, p. 30). In the case of the diagnostic test in its pilot version, it was not possible to ensure that the pupils taking the diagnostic test also took another test serving as a criterion measure. This was mainly due to practical reasons, such as the wide choice of available and convenient standardised tests, the necessary parental consent to testing, and the financial costs.

Predictive validity "measures how well a test predicts performance on an external criterion" (Davies et al., 1999, p. 149). Working with predictive validity in the case of the diagnostic test was difficult, as a large number of factors other than language (e.g., subject knowledge, intelligence, motivation, etc.) came into play. However, it would be possible to ask teachers directly for feedback if special lessons were provided to children of migrants at school, and/or if the particular pupil was included in a group learning in Czech (taking the results of the diagnostic test into account), or if there was highly modified teaching of the second language based on the diagnostic test. Unfortunately, this feedback would, to a certain extent, be subjective and based on the untrained judgements of supervisors.

Another possibility for investigating construct validity is through think-aloud protocols and/or retrospections. However, this method did not seem to be practical due to the age of the respondents and the time required.

It is obvious that a system through which predictive validity can be verified needs to be introduced.

Reliability

Reliability in Reading and Listening

Analysing the data gained from pretesting led to verifying whether, and how, the tasks function, and to calculating reliability coefficients. For both diagnostic tests, we used the statistical software Iteman 4.1, based on Classical Test Theory. In the case of the lower-grader diagnostic test comprising A1 and A2 levels, the tasks analysed were the Reading and Listening tasks at both levels and the first Writing task at level A1. The test was taken by 129 respondents. In the case of the upper-grader diagnostic test comprising A1, A2 and B1 levels,

Reading and Listening at all levels and the first A1 task in Writing⁷ were analysed. The test was taken by 132 respondents.

Reliability of a test can be estimated in two ways: by parallel measurements (test-retest method, parallel test method) or by internal consistency (splitting the test into two halves and estimating the internal consistency). For the test-retest method, it is necessary to re-take the test after a certain period of time. This method was considered unfeasible in the case of the diagnostic tests in question because it would require testing the same pupils after some time. It proved difficult to gather the same test-takers again and/or gain their and their parents' consent for retaking the test. Using parallel tests was not considered practical either, as there would have to be two parallel versions of the test and pupils would have to take both of them, which would be demanding and time consuming, especially considering the children's age.

The most frequently used method of estimating reliability is the internal consistency method, which can only be applied to tests with homogenous content. This method presupposes that the answers to all items measuring the same characteristics hold sufficiently high positive correlation, and that if the test is reliable, its parts – its two halves – must also be reliable. These halves are assessed separately and then the results are correlated. The correlation between the two halves is corrected using the Spearman-Brown Formula (Chráska, 2007).

Table 3 shows the reliability coefficients gained by applying the Kuder-Richardson Formula in the lower-grader test as a whole, as well as in its two parts. It also shows the reliability coefficient gained by the Split Half method in three variants of halving the set: Split-Half Random (items are split into halves at random), Split Half First-Last (one set consists of the first half of the items, the other set of the second half), and Split Half Odd Even (one set comprises the odd items, the other one the even items). For all of the variants of splitting, the results are shown for both non-corrected variants and the variants corrected by the Spearman-Brown Formula. This correction is used because in the non-corrected version we compare two tests with only half of the items contained in the live test. Standard error of measurement (SEM), which estimates the standard deviation of the errors of measurement in the scale scores, is also reported. Regarding the values of the reliability coefficient, Chráska (2007) claims that a reliability coefficient of 0.8 and above is generally considered optimal for didactic tests, while 0.95 is excellent.

7 This task consists of five questions about the pupil, usually requiring a one-word answer, with the responses being rated for content only, i.e., whether the pupil answers the question or not, which makes such an analysis possible.

Table 3. *Reliability coefficients for the lower-grader diagnostic test*

	Alfa (KR-20)	SEM	Non-corrected			Spearman-Brown Correction		
			Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)
Whole test	0.951	2.180	0.921	0.796	0.919	0.959	0.886	0.958
A1 test	0.915	1.373	0.844	0.755	0.878	0.915	0.860	0.935
A2 test	0.915	1.663	0.804	0.794	0.867	0.891	0.885	0.929

The data in Table 3 show that when applying the Kuder-Richards formula the reliability coefficient exceeds 0.9 for the individual tests and even reaches 0.95 for the whole test. Slightly lower reliability coefficients occur when using the split-half method. However, it should be noted that splitting a diagnostic test in two equivalent halves is complicated. Since the tasks and items are ordered according to their difficulty, we get the lowest reliability coefficient when comparing the first and the second half of items (Split-Half First-Last Method). The reliability coefficient is considerably higher when the Random or Odd-Even variant of the Split-Half method is used. In these cases, it almost always exceeds 0.9.

Similarly to Table 3, Table 4 shows the same test characteristics for the upper-grader diagnostic test.

Table 4. *Reliability coefficients for the upper-grader diagnostic test*

	Alfa (KR-20)	SEM	Non-corrected			Spearman-Brown Correction		
			Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)
Whole test	0.971	2.523	0.904	0.798	0.952	0.949	0.888	0.976
A1 test	0.920	1.319	0.825	0.697	0.878	0.904	0.821	0.935
A2 test	0.944	1.190	0.898	0.769	0.921	0.946	0.869	0.959
B1 test	0.934	1.663	0.881	0.805	0.885	0.937	0.892	0.939

In this case, when the Kuder-Richardson Formula is applied the reliability coefficients are even higher than in the case of the lower-grader test. High values of the reliability coefficient are also gained when using the Split-Half method.

Ensuring reliability of scoring written and spoken performances

Written and spoken performances are assessed on the basis of detailed written criteria. All performances were assessed by one of two experienced raters trained to use the scale. Questionable performances were discussed by both raters, but as double marking was not introduced, inter-rater reliability has not been counted. In Speaking, one of the raters acted as an interlocutor, the other as a rater.

Impact

Impact is traditionally perceived as “the effect of a test on individuals, on the educational system and on society in general” (Davies et al., 1999, p. 79). A more detailed study dealing with the influence of the diagnostic test on pupils and the teaching process is still to be undertaken. However, it is already possible to consider whether this test can help to solve certain issues that teachers pointed out during the qualitative research for the project.⁸ Among other things, the research showed the following:

- The practice of accepting children of migrants at Czech schools differs considerably.
- The decision about which grade the pupil should attend is usually made at a meeting between the school principal, the class teacher and the Czech language teacher (teaching Czech as the first language).
- The main criteria for placing children of migrants in particular grades include the age of the child, her/his L1, her/his current level of communicative competence in Czech, and the results of the child’s last school report. Other criteria can also be taken into account. These would typically include the possibilities available at the school and among its pedagogical staff (e.g., the class teacher’s knowledge of foreign languages, her/his personality, the number of pupils in the class, the number of pupils with L1 other than Czech, and the final composition of nationalities in the class). The tendency to place the pupils in grades primarily according to their age, not according to their level of communicative competence in Czech, was dominant.
- As mentioned above, the pupil’s level of communicative competence also played a role when deciding which grade the pupil should attend, although it was not the most important factor. However, it should be noted that there was no unified, standardised way of testing: language skills were assessed more or less intuitively.

8 More detailed results of this research can be found in Kostecká et al. (2013).

- Some pedagogues do not realise that it is not only possible but even advisable to take into account the level of communicative competence in Czech in subjects other than just Czech Language and Literature when assessing pupils with L1 other than Czech.
- The activities aimed at supporting the integration of pupils with L1 other than Czech are determined, in part, by the financial and human resources of the school. These activities may include preparatory classes, intensive summer courses and placing the pupil directly in common classes while also assigning her/him an assistant who can teach the pupil Czech intensively, supplementing the pupil's attendance at language courses throughout the school year.

We assume that the diagnostic test for pupils with L1 other than Czech would have a positive impact on a number of the points listed above. However, this assumption must be supported by further research, designed similarly to the qualitative research conducted prior to launching the diagnostic test, but this time focusing on changes brought about by the implementation of the test.

Practicality

One of the fundamental features of the test is practicality, as “however valid and reliable a test may be, if it is not practical to administer it in a specific context then it will not be taken up in that context” (Davies et al., 1999, p. 148). In the case of the introduced diagnostic test, practicality relates in particular to the following areas:

- The length of the test (respecting at least the minimum number of items that are essential in order to consider the test reliable, while at the same time taking into account the attention span of the given age group and the total time allotted to complete the test).
- The order of the subtests (Listening had to be placed as the first subtest so that the pupils could continue at their own pace).
- The demands related to the administration of the test, so that the administration can be left to trained staff at the school if necessary.
- The demands related to prompt rating, so that it is clear whether the pupil should take the diagnostic test on a higher level.
- The demands related to rating, so that at least the receptive skills can be assessed directly at schools by trained raters, not by an external team of specialists.
- The financial costs of maintenance of the test, which represent one of the points of current interest.

Conclusions

In Parts 3 and 4, we not only demonstrated what was done with regard diagnostic testing within the project, but we also identified fields in which further development is desirable.

Firstly, maintenance of the diagnostic test should be assured. This concerns not only financial support, but also human resources.

Secondly, it may be necessary to train administrators, examiners and possibly also raters if the number of test takers grows. In the piloting and pre-testing phase, these roles were able to be handled by the team of test constructors, as the number of test takers was relatively low. If the test is used on the national level (although probably voluntarily), more staff will be required to participate in test administration, examination and assessment. In diagnostic testing, prompt and detailed feedback both to test takers and teachers or schools is crucial. With growing numbers of test takers, it may also be necessary to train a number of experts in providing feedback to the test users.

Thirdly, it is worth exploring the impact of the diagnostic test on teaching, pupils and teachers, as well as on schools. Although the feedback in this regard might be limited, due to the fact that the diagnostic test has not yet been introduced on a national basis, it is obvious that it would serve as valuable material and would verify whether the diagnostic test has been used in accordance with the intentions of the test developers.

Given the rapid increase in immigration to the Czech Republic in the past 20 years, the educational integration of pupils who are not native speakers of Czech is a subject that is a very relevant issue today. Since the number of immigrants in the Czech population is likely to grow even more, its relevance will only increase in the future.

The diagnostic test for lower graders and upper graders at Czech primary schools whose L1 is different from the language of instruction represents one of the first attempts to design an instrument that would help teachers, schools and children of migrants with (language) integration.

Acknowledgements

This article was written with the support of the Czech Science Foundation, within the project entitled “Integration of Children of Non-Nationals from Czech Lower Secondary Schools”, registration number: 13-32373S.

Reference

- Alderson, J. Ch. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J. Ch., Brunfaut, T., & Harding, L. (2015). Towards a theory in diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), 236–260.
- Alderson, J. Ch., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Cvejnová, J., Papírníková, I., & Benešová Grunotvá, D. (2007). *Diagnostika úrovně znalosti českého jazyka (diagnostická příručka)* [Diagnosing the Level of Czech]. Praha: Centrum pro integraci cizinců.
- Davies, A. et al. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Drbohlav, D. (2011). Imigrace a integrace cizinců v Česku. Několik zastavení na cestě země v její migrační proměně z Davida na téměř Goliáše [Immigration and Integration in the Czech Republic]. *Geografie*, 116(4), 401–421.
- Nancy, F., & Alba, R. (2008). Immigrant Religion in the U.S. and Western Europe: Bridge or Barrier to Inclusion? *International Migration Review*, 42(2), 360–392.
- Harding, L., Alderson, J. Ch., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 32(3), 317–336.
- Heckmann, F., & Schnapper, D. (2003). *The integration of immigrants in European societies: National differences and trends of convergence*. Stuttgart: Lucius & Lucius.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: a quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1–11.
- Chráška, M. (2007). *Metody pedagogického výzkumu: základy kvantitativního výzkumu* [Methods of Pedagogical Research: A Basis for Quantitative Research]. Praha: Grada.
- Ioannou-Georgiou, S., & Pavlou, P. (2013). *Assessing young learners*. Oxford: Oxford University Press.
- Janská, E., Průšvicová, A., & Čermák, Z. (2011). Možnosti výzkumu integrace dětí Vietnamců v Česku: příklad základní školy Praha-Kunratice [Possibilities of Researching the Integration of Vietnamese Children in the Czech Republic]. *Geografie: sborník České geografické společnosti*, 114(4), 480–496.
- Kostecká, Y. et al. (2013). *Žáci-cizinci v základních školách: Fakta, analýzy, diagnostika* [Pupils-Foreigners at Primary Schools: Facts, Analyses, Diagnostics]. Praha: Univerzita Karlova v Praze, Pedagogická fakulta.
- McKay, P. (2010). *Assessing young learners*. Cambridge: Cambridge University Press.
- Nováková, S. et al. (2001). *Evropské jazykové portfolio pro žáky do 11 let v České republice* [The European Language Portfolio for Pupils up to 11 Years in the Czech Republic]. Plzeň: Fraus.

Perclová, R., & Marešová, E. (2001). *Evropské jazykové portfolio pro žáky a žákyně ve věku 11 – 15 let v České republice* [The European Language Portfolio for Pupils Aged 11–15 in the Czech Republic].

Praha: Fortuna.

Suárez-Orozco, M. M. (Ed.) (2007). *Learning in the global era: International perspectives on globalization and education*. Los Angeles: University of California Press.

Šindelářová, J., & Škodová, S. (2013). *Čeština jako cílový jazyk 1. díl: Metodika pro výuku osob (dětí do 15 let) s mezinárodní ochranou formou azylu nebo doplňkové ochrany* [Czech as a Target Language, Part 1]. Praha.

Biographical note

KATEŘINA VODIČKOVÁ graduated in Czech and English Studies, received her Ph.D. in Czech Language and her MA degree in Language Testing at the Lancaster University, UK. She has worked for Charles University since 2002. She gives lectures, teaches Czech and has been involved in examination development in a number of projects.

YVONA KOSTELECKÁ graduated from the Charles University, Faculty of Science and obtained Ph.D. in Education. She works as the head of Institute for Professional Development, Faculty of Education, Charles University. Her research focuses on integrating foreigners into the Czech school system, and on international migration of researchers and skilled workers.