Ukrainian language proficiency test review

Daniil M. Ozernyi ⓘD [GQ: 1]
Northwestern University, USA

Ruslan Suvorov ⓘD
University of Western Ontario, Canada

Keywords
CEFR, test fairness and justice framework, test review, Ukrainian, validity

General description [AQ: 1]

The Ukrainian Language Proficiency (ULP) test, officially titled *Exam of the level of mastery of the official language* (*Ispyt na riven' volodinnya derzhavnoyu movoyu*) is a new test launched in Summer 2021. The name of the test in Ukrainian, incidentally, does not contain the words "Ukrainian" or "foreign language." According to the state regulations (Kabinet Ministriv Ukrayiny [KMU], 2021a; Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy [NKSDM], 2021a, 2021b), the levels of mastery of Ukrainian in the test are aligned with the CEFR levels.[1] The test was introduced as a product of the law about the official language of Ukraine which mandated that civil servants and citizens who are being naturalized are fully able to use Ukrainian in performing their duties. The mentioned law came, in turn, as a measure to ensure that Ukrainian language and heritage survive against the assault from Russia that has been present since 1617 and flared up in 2014, culminating in a full-scale ongoing genocide of Ukrainians as of 2022.

The ULP test comprises two versions: (a) ULP for acquisition of Ukrainian citizenship (*Ispyt na riven' volodinnya derzhavnoyu movoyu (dlya nabuttya hromadyanstva)*), and (b) ULP 2.0 for holding civil office (*Ispyt na riven' volodinnya derzhavnoyu movoyu 2.0 (dlya vykonannya sluzhbovyh obov'yazkiv)*). To differentiate between the two versions of the test in this review, we will refer to the former version as ULP-C and to the latter version as ULP 2.0.

The purpose of this review is to apply Kunnan's (2018) fairness and justice framework to evaluate both ULP-C and ULP 2.0 since they are united by (a) the alignment

Corresponding author:
Daniil M. Ozernyi, Department of Linguistics, Northwestern University, 2332 Campus Drive, Evanston, IL 60201, USA.
Email: doz@u.northwestern.edu

with the CEFR scale which poses ULP 2.0 as a continuation of ULP-C, (b) the same raters and administrators, and (c) the common recently-developed descriptors for B1, C1, and C2 levels of Ukrainian.

Author and publisher

The ULP test was developed and is currently administered by the National Committee on the Standards of the Official Language (NKSDM), which is a state agency that is part of the Ministry of Education and Science in Ukraine.

De jure address: 10 Peremohy prospekt, Kyiv, Ukraine 01135
Postal address: 4 Nestorivskyi provulok, Kyiv, Ukraine 04053
Press service phone: +38 (044) 235-00-64
Website: mova.gov.ua

Test purpose and context

The two versions of the test were created with two primary objectives: (a) to ensure that those coming to Ukraine are proficient in Ukrainian, the sole official language (ULP-C); and (b) to enforce the state law (see Vidomosti Verhovnoyi Rady [VVR], 2019) requiring that all persons holding civil office have sufficient mastery of Ukrainian to perform their duties (ULP 2.0). The latter objective is necessitated by some residual dominance of the Russian language, which, while steadily dwindling down, still persists in some regions of Ukraine and can affect the Ukrainian language proficiency of individuals who reside there (Ozernyi, 2021). Among these individuals, some speak regional vernacular dialects (known in Ukrainian as "surzhyk") that fuse linguistic features of both Ukrainian and Russian, whereas others use Russian as their dominant language. This latter segment of population comprises adults who received their education more than 30 years ago in Soviet Ukraine when the language of instruction was Russian. Considering the afore-mentioned segments of population, ULP 2.0 was introduced to ensure that everyone who holds civil office possesses adequate proficiency in Ukrainian to perform their duties.

Target audiences for ULP-C and ULP 2.0

As ULP-C is designed for those who are seeking Ukrainian citizenship, the majority of the target population are likely to be speakers of Ukrainian as a foreign language, with only a small fraction of test takers being heritage speakers or other kinds of bilinguals. Unlike ULP-C that purports to measure Ukrainian language proficiency among foreigners who are non-native speakers, ULP 2.0 is designed for Ukrainian citizens seeking to hold public office who are primary speakers of a language other than Ukrainian (typically Russian), but who have some degree of proficiency in Ukrainian as they have been exposed to Ukrainian for most of their lives. Although the target audiences for ULP-C and ULP 2.0 might not seem all that different at first glance as they both have limited ability to speak and write in Ukrainian,[2] the differences among them are in fact quite stark, which has significant implications for the construct measured by ULP 2.0.

ULP 2.0's target audience comprises receptive bilinguals who have near-native receptive skills, but whose ability to produce language is either absent or inadequate (Sherkina-Lieber et al., 2011). As a result, ULP 2.0 should be viewed not as a test for foreign language learners, but as a particular language test tailored to specific needs of the vast population of receptive bilinguals.

Length and administration

The length of ULP-C is 30 min, whereas the length of ULP 2.0 is 150 minutes (KMU, 2021b).[3] Although both versions of the test are regulated by NKSDM, they are administered through a network of authorized institutions, which are almost exclusively Ukrainian universities. ULP is a computer-based test with an option for special accommodations. The test results are made available through the NKSDM's website (*mova. gov.ua*) no later than 15 days after the day of administration (KMU, 2021a). The results are reported as a percentage score without any points for individual sections or breakdown of mistakes.

Structure, task types, and points allocation

ULP-C consists of listening (12 points), reading (31 points), writing (33 points), and speaking (38 points) parts, resulting in a total of 114 possible points. Despite these parts containing different numbers of points, each part is weighted equally and worth 25% of the total score. The listening part comprises 12 multiple-choice and true-false listening items, whereas the reading part consists of 31 matching, multiple-choice, and true-false reading items. Next, the writing part contains two writing tasks: a 50-word written message (such as an invitation or a congratulatory note) and a 100-word essay on a given topic. Finally, the speaking part has three tasks: a picture description task (about 10 sentences[4]), a monologic discussion of a given topic (20–25 sentences), and a dialogue with the examiner on a given topic (about 5 sentences; see examples in NKSDM, 2022a).

The structure of ULP 2.0 is somewhat different, consisting of a combined "language culture, writing, and reading" part (44 points, or 70% of the total score) and a speaking part (19 points, or 30% of the total score). The combined "language culture, writing, and reading" part includes selected-response item types (such as multiple-choice items, true/false items, and matching items) and a writing task that requires test takers to listen to an audio and produce a 100-word written response that summarizes the content of the audio and shares their opinion about the topic. The speaking part of ULP 2.0 comprises two tasks: a dialogue with the examiner about a given situation (7–10 sentences) and a monologue detailing the test taker's perspective on one of the given topics (15–20 sentences, see examples in NKSDM, 2022b).

To pass ULP-C, test takers must attain the B1 level (at least 50%), whereas for ULP 2.0 the passing level is C1 (at least 70%, see Table 1). Table 1 shows the percentage allocation for the levels and the outcomes for each version of the test. It should be noted that NKSDM does not specify whether ULP-C actually assesses the levels above B1. Although the presence of C1/C2-level descriptors (e.g., NKSDM, 2021a, 2021c) implies

Table 1. Points allocation in ULP-C and ULP 2.0.

Levels		Percentage	Outcomes for ULP-C	Outcomes for ULP 2.0
Beginner level	A1	30–39	Fail, no certificate.	Fail, A1 certificate
(*Pochatkovyi riven'*)	A2	40–49		Fail, A2 certificate
Intermediate level	B1	50–59	Pass, B1 certificate.	Fail, B1 certificate
(*Seredniy riven'*)	B2	60–69	Pass, B1 certificate. The same	Fail, B2 certificate
Proficient level	C1	70–89	certificate is given for any	Pass, C1 certificate
(*Vil'ne volodinnya*)	C2	90–100	score between 60 and 100.	Pass, C2 certificate

ULP: Ukrainian Language Proficiency.

this being the case, the statutes (i.e., NKSDM, 2021b) suggest that all test takers receiving 50% to 100% for ULP-C get the same B1-level certificate. The lack of any explicit assessment of levels above B1 is most likely due to the short 30-minute duration of ULP-C.

Price

ULP 2.0 is free for all eligible test takers, whereas ULP-C costs 492 UAH (approximately US$14 as of July 2022) for all test takers except for refugees who apply for Ukrainian citizenship (KMU, 2021a).

Appraisal of the ULP test

Our appraisal of the ULP test is guided by Kunnan's (2018) fairness and justice framework. In brief, this framework is rooted in Toulmin's (1958, cited in Kunnan, 2018) model of argumentation, according to which a sound argument should comprise a claim that is based on grounds (i.e., foundation or an underlying principle of a claim) and linked to it by a warrant (i.e., elaboration of a claim) that requires backing (i.e., evidence to support the warrant). The two main principles or grounds undergirding Kunnan's (2018) framework are fairness and justice that can be linked to a series of claims, with a claim being defined as "an assertion . . . made by an assessment agency or developer . . . that informs the public as to what ability is being assessed . . . and for what purpose" (Kunnan, 2018, p. 91). According to Kunnan (2018), claims should be based on ethical principles and supported by empirical evidence that serves as backing. In our appraisal of the ULP test, we review some of the key claims made by NKSDM and evaluate the extent to which these claims are supported by the available evidence.

Claim 1: The ULP is a meaningful (valid) test as it is aligned with the CEFR

Meaningfulness, the term used by Kunnan (2018) to refer to validity, is arguably one of the most essential principles of language assessment. The first claim concerns the alignment of the ULP with the CEFR that NKSDM positions as an indication of the

meaningfulness of the ULP. In particular, NKSDM claims that the levels of mastery in the ULP test are aligned with the CEFR levels; however, there are only two documents available that explain the levels used in the test: NKSDM (2021a, 2021c). NKSDM (2021c) provides a general overview of the six levels of mastery used in the ULP test, whereas NKSDM (2021a) lists all the descriptors for B1, C1, and C2 levels, but only for ULP-C and not for ULP 2.0. We did not find any evidence of descriptors for A1, A2, and B2 levels for either version of the test. Using the available documents, we reviewed two types of alignment in relation to this claim: (a) the alignment between the NKSDM's (2021c) classification of the mastery levels and the summary of the CEFR global scale levels (Council of Europe, 2001), and (b) the alignment between the NKSDM (2021a) descriptors for B1, C1, and C2 levels and the CEFR updated descriptors for those three levels (Council of Europe, 2020).

With regard to the first type of alignment, the NKSDM's (2021c) classification of the six levels is a verbatim translation of the CEFR global scale levels (Council of Europe, 2001, p. 24), suggesting the direct alignment between the two sets of levels.

When analyzing the second type of alignment, we identified a number of issues and discrepancies. First, while all the B1, C1, and C2-level descriptors outlined in NKSDM (2021a) are written as can-do statements (which is in line with the CEFR), they are grouped into skill-based categories of listening, reading, writing, and speaking, as well as the "communicative goals" (e.g., language functions such as "asking for help") category. This traditional four-skills approach is different from the communicative approach used by the CEFR that organizes all descriptors around communicative language competences (i.e., linguistic, sociolinguistic, and pragmatic), activities (i.e., reception, production, interaction, and mediation), and strategies (see Council of Europe, 2020, p. 32). Second, unlike the CEFR, NKSDM does not appear to differentiate between C1 and C2 levels as all the descriptors for these two levels outlined in NKSDM (2021a) are identical. Unfortunately, we were not able to find any information explaining the NKSDM's rationale for using two different levels (C1 and C2) with identical descriptors. Third, the majority of NKSDM's (2021a) level-specific descriptors are generic, lack qualifiers, and appear to be overly simplified versions of the same-level descriptors in the CEFR. For example, one of the NKSDM's (2021a) B1-level descriptors in the "Speaking" section states that a learner at this level "can orally describe people, places, objects, events, and actions," whereas similar B1-level descriptors in the CEFR contain specific qualifiers (e.g., "can *reasonably fluently* sustain a straightforward description of one of a variety of subjects . . .") and linguistic functions and forms (e.g., "can *express opinions* on subjects relating to everyday life, *using simple expressions"*) that allow for a more nuanced definition of the language abilities representative of that level. Finally, we identified several cases when B-level descriptors from the CEFR were presented as C1/C2-level descriptors in NKSDM (2021a), as well as instances of using identical descriptors for B1, C1, and C2 levels in NKSDM (2021a) (e.g., "can participate in a dialogue").

We have reached out to NKSDM asking whether there was any evidence to support ULP-C and ULP 2.0's alignment with the CEFR that was not available publicly. NKSDM referred us to NKSDM (2021a) and three papers by Mazuryk, one of the NKSDM committee members, published in 2006–2009; however, none of these documents contained any empirical evidence supporting the alignment of ULP-C or ULP 2.0 with the CEFR.

For instance, the paper by Mazuryk (2009) claimed that the six levels of Ukrainian proficiency used in the local test of Ukrainian (administered at Lviv National University of Ivan Franko) were based on the CEFR levels; however, the relationship between this locally administered test of Ukrainian and ULP-C or ULP 2.0 was not established or explained. When we asked NKSDM whether there was any documentation supporting the alignment of the ULP tests with the CEFR descriptors (Council of Europe, 2020), the agency referred us to Mazuryk et al. (2018) who provide a detailed overview of the six levels and descriptors for Ukrainian, aver that they are aligned with the CEFR levels and descriptors, but, unfortunately, yield no empirical evidence for this alignment.

Finally, some of the ULP tasks do not seem to be capable of distinguishing between the CEFR levels that the test purports to measure. For example, the length of the writing task for ULP 2.0 is described as "not fewer than 100 words" in NKSDM (2021e).[5] Because ULP 2.0 aims at measuring higher levels of CEFR, such short writing samples appear to be inadequate for evaluating the many elements of students' written production listed in the 2020 CEFR Companion Volume. For example, descriptors for overall written production stipulate that at C1 learners can "write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion" (CEFR, 2020, p. 75), which cannot be adequately accomplished in a 100-word text.[6] **[AQ: 2]**

To conclude, backing in support of the alignment between the ULP and the CEFR appears to be largely limited and, as a result, Claim 1 remains largely unsupported, necessitating further empirical evidence.

Claim 2: The ULP is meaningful in terms of its construct and tasks

According to Kunnan (2018), claims about the meaningfulness (validity) of a test can be supported by evidence from the analysis of the test content and construct. Specifically, such evidence can come from the analysis of test specifications and the actual test items and tasks. To seek support for the second claim, we analyzed language-specific syllabuses for ULP-C (NKSDM, 2021f) and ULP 2.0 (NKSDM, 2021e). These syllabuses (called "prohramas") are intended to be used as test preparation guidelines that outline the test structure, content, types of tasks, and language-specific competences that test takers are expected to demonstrate when completing a specific version of the ULP.

The syllabus for ULP-C states that this version of the test is designed to measure communicative language competences and language activities and strategies such as reception, production, interaction, and mediation at the B1 level (NKSDM, 2021f). However, the rest of the syllabus is organized primarily around the four skills (i.e., listening, speaking, reading, and writing) and the description of what test takers are expected to do at the B1 level in each skill. Although the use of the four-skills model may not necessarily be problematic, we did not find any documentation explaining the connection between the four skills and the communicative language competences and strategies that the test purportedly measures, such as plurilingual competence, which can be particularly relevant for Ukrainian–Russian bilinguals. In other words, there appears to be a mismatch between the *holistic approach* to measuring overall language proficiency (as a set of

communicative language competences and strategies) that is reported for ULP-C (and adopted by the CEFR) and the *traditional four-skills approach* used in the test preparation materials (i.e., syllabuses) and the actual test. Although this might be the case for most of the language tests on the market, ULP-C does not account or acknowledge such a mismatch.

The analysis of the syllabus for ULP-C also reveals several points of ambiguity. First, when describing specifications for the listening section of ULP-C, the syllabus states that "the duration of the authentic audio recordings [is] up to 2 minutes" (note the plural form of the second noun), the transcript can contain up to 400 words, the percentage of new words is up to 3%, and the speech rate is 120 words per minute (NKSDM, 2021f, p. 2). This statement, however, can be interpreted in two ways: (a) that the entire listening comprehension section is up to 2 minutes long and comprises a 400-word audio-recording, or, alternatively, (b) that there is a number of audio-recordings, each of which can be up to 2 minutes long, with the total length of the transcript for all of them being up to 400 words. With regard to the first interpretation, if the speech rate is 120 words per minute, then a single 2-minute audio text in the listening section would contain only up to 240 words, not 400 words indicated in the syllabus. If we use the second interpretation and assume that a 2-minute length applies to each audio-recording, then the use of a 400-word transcript with the speech rate of 120 words per minute suggests that the entire listening section contains only two audio-recordings with the total listening time of only 3–3.5 min, which appears to be quite short for effectively measuring listening comprehension skills at the B1 level.[7] Upon our request, NKSDM provided us with a sample script for one of the test questions that was 50 words long. It is unclear how this sample aligns with what is stated in the syllabus for the test as the sample is neither 400 words long nor lasting 2 min.

Second, the section of the syllabus that describes the requirements for lexical knowledge states that "the lexical knowledge should comprise at least 2500 words and collocations [and] the active lexical knowledge should comprise at least 1200 words and collocations" (NKSDM, 2021f, p. 5). These numbers do not appear to be justified and seem entirely arbitrary. More importantly, the wording used in the above-mentioned statement does not specify the type of lexical units (e.g., word types, word families, lemmas, or flemmas) that the test takers are expected to demonstrate. With Ukrainian being a richly inflected language, the knowledge of 2500-word families, for instance, would require a much higher proficiency than the knowledge of 2500-word types.

Finally, the most intriguing part of the syllabus for ULP-C is the section titled "The contents of language material" which contains a plain list of grammatical topics, for example, "Adjective. The meaning of adjectives. [. . .] Possessive adjectives. Conjugation of adjectives (strong and weak groups). Degrees of comparison. The syntactic connection of adjectives and nouns" (NKSDM, 2021f, p. 4). This list is highly reminiscent of knowledge relating to the so-called metalinguistic competence (*sensu* Falk et al., 2015). However, the language proficiency cannot be adequately determined by measuring the knowledge or awareness of concepts such as "possessive adjectives" or "strong and weak declensions" (cf. Epstein et al., 1996).

The syllabus for ULP 2.0 (NKSDM, 2021e) is markedly similar to that of ULP-C, except for a tripled list of metalinguistic concepts, including "types of [phonetic] errors

and ways of correcting them" (p. 7); "metaphor as a means of semantic derivation" (p. 7); "euphemism, periphrasis, amplification, pleonasm" (p. 7); and "syntactic features of the infinitive" (p. 11). The relevance of these metalinguistic concepts to the language proficiency and ability to communicate in Ukrainian at the C1–C2 levels, is unclear.

Interestingly, the transcript length in the listening section of ULP 2.0 stays the same for the C1 level (i.e., up to 400 words), while the speech rate is left unspecified and the length of the audio is extended to 4 min. Unfortunately, we were unable to find any documentation that would provide justification for these design choices or any reliability estimates (although the fact that the test contains only 12 listening items suggests that the reliability of the listening section should be fairly low). In sum, many of the test specifications lack a theoretical and/or empirical basis, and there appear to be inconsistencies between documented specifications and actual test content.

To conclude, evidence in support of Claim 2 appears to be tenuous. Contrary to research-driven specifications for other language exams (e.g., the English Profile Program; Kurteš & Saville, 2008 *et seq.*), the specifications outlined in the syllabuses for both versions of the ULP test suggest that many test design choices appear to be arbitrary rather than informed by theory or research, thus undermining the validity of interpretations that can be made on the basis of test scores from ULP-C and ULP 2.0.

Claim 3: The ULP is fair to all test takers as it ensures a reliable scoring procedure

With transparency and uniformity being two of the four core aspects of fairness and justice in Kunnan's (2018) framework, we also sought evidence for the claim that the ULP is fair to all test takers as it ensures a reliable scoring procedure.

In terms of transparency, ULP is a markedly fair test: It appears that all of the information about the process of creating the test is made public, as it should be when a government agency is tasked with authoring and administering the test. It is important to convey the extent to which this statement is true: The Ukrainian law mandates that every document (except for classified ones) issued by the committee be available to any Ukrainian citizen. As a result, every test takers' certificate with the ULP results is publicly available on the NKSDM's website. The uniformity of transparency, however, is not ensured as far as rater selection and rater training are concerned. NKSDM effectively uses the same people to score ULP-C and ULP 2.0. Furthermore, the only requirement to become a rater is having at least a Master's degree in Ukrainian language and literature (NKSDM, 2021d) and, oddly, being a faculty member of a university which administers the exam. Intriguingly, no experience with second language learning, teaching, or assessment is required.

Furthermore, we contacted NKSDM asking them to provide documentation for training procedures of the raters. NKSDM directed us to a syllabus for rater training that prescribes 15 hours of training, six of which are lectures, four are practice, and five are self-study. Description of the content prescribed for every of the 15 hours was also provided. However, upon careful scrutiny, we found that only four out of 15 hours actually involved discussing the rubrics for writing/speaking of both exams, whereas the remaining 11 hours were devoted to tasks such as reading the relevant statutory acts, learning

the history of creating NKSDM and personalia of its current members, and using the online platform to submit scores. Crucially, no hands-on rating experience could be found in the syllabus for rater training. NKSDM also clarified that they have been training raters since July 2021 and currently have 259 raters. Furthermore, the agency stated that, as of July 2022, they have administered 125,388 ULP 2.0 tests and 627 ULP-C tests for a total of 126,015 test administrations. Even if we assume that NKSDM had 259 raters all along[8] and that each test taker's performance was evaluated only by one rater (which is clearly not a good practice but which is stated in KMU, 2021a), it appears that every rater evaluated on average 486 test takers in a year or less.

In addition, the scoring of test takers' open-ended written and oral responses on the ULP test appears to be based on a list of general evaluative categories and score ranges for each category rather than on validated analytic rubrics with specific descriptors (see Section 4 in NKSDM, 2021b). According to NKSDM (2021b), for example, the category "lexical knowledge" can be assigned 0–4 points; however, there is no indication as to which criteria a rater should use to assign a specific number of points. Without having an analytic rubric with level-specific descriptors, raters seem to be expected to use their subjective judgment. (Note that when we asked NKSDM about the rubric, they provided us with the same generic rubric that is available online.) It is possible that the lack of specific descriptors or evaluation criteria may have a detrimental impact on inter- and/or intra-rater reliability and, consequently, undermine the fundamental aspects of fairness and justice in the ULP test. Unfortunately, we did not find any evidence of research on rater reliability for the ULP test that would demonstrate the consistency of raters' scoring.

In sum, while we recognize and applaud the NKSDM's commitment to openness, we find Claim 3 to be largely unsupported due to the lack of empirical evidence demonstrating the reliability of the scoring procedure and the fact that each test taker's performance is evaluated only by one rater.

Conclusion

Being the first official language test designed, developed, and endorsed by the government of Ukraine, the Ukrainian Language Proficiency test is undoubtedly an important and welcome step toward creating a centralized mechanism for measuring proficiency in this less commonly taught (and assessed) language in the global context. However, as this review demonstrates, there are a plethora of issues surrounding this test that need to be resolved before it can be justifiably used for large-scale standardized assessment. Although it is laudable that NKSDM is open to communication and makes the entirety of the ULP's documentation available to the public, there appears to be a lack of validation research to support the claims made by the developers and administrators regarding the test. Furthermore, our review of existing documentation revealed insufficient clarity about the test construct, multiple discrepancies between syllabi and test samples, and the lack of specific descriptors in the rating rubrics. In addition to the mismatch between the reported purpose and the actual content of the test—and contrary to the developers' claims—we found little evidence that the ULP is aligned with the CEFR as the ULP's points of intersection with the CEFR seem to be few and sporadic. Crucially, it appears

that the test is not built on teacher-driven or experience-driven discussions about Ukrainian as second/foreign language or corpus data. Instead, it is based on a seemingly arbitrary divide of a list of grammar rules of Ukrainian into six sections, as written by researchers and activists who study Ukrainian language and literature, not first-/second-language teaching or language acquisition. As of now, validity seems to be reduced to the NKSDM's claim about the purported alignment with CEFR—which remains mostly unsubstantiated, as suggested by this review. Although the limited validity evidence for the ULP can be attributed to the lack of validation research on this test, it is also necessary to acknowledge the lack of resources available for developing a test of Ukrainian, such as the absence of comprehensive native and/or learner corpora for Ukrainian and the dearth of research on acquisition of Ukrainian as L2. We suggest that NKSDM follow the steps of T-series (van Ek, 1975 et seq.) which was one of the earliest attempts at structured inquiry into the stages of acquiring English as L2. We also hope that NKSDM will consider inviting experts in second language acquisition, testing and assessment to conduct validation research to support the aforementioned claims about the test and make revisions necessary for improving the validity of the interpretations and inferences that can be made on the basis of the ULP scores.

ORCID iDs[GQ: 3]

Daniil M. Ozernyi ⓘD https://orcid.org/0000-0001-6998-9855
Ruslan Suvorov ⓘD https://orcid.org/0000-0003-2736-5205

Notes

1. Although there was a legislative attempt (see Ministerstvo Osvity i Nauky Ukrayiny [MONU], 2018) to create a framework of levels based on the CEFR, the ULP does not make reference to that legislature. Instead, both NKSDM (2021a) and NKSDM (2021e, 2021f) reference the CEFR (Council of Europe, 2001, 2020) directly.
2. Note that there is a notable mismatch between what ULP 2.0 intends to measure (i.e., the ability to produce formal spoken and written discourse in Ukrainian when performing official duties as civil servants) and what it appears to measure. Yet, NKSDM clearly states that the test is designed to assess "general language proficiency" and does not select aspects of the language relevant to performing the above-mentioned duties.

3. We found a reference that seems to indicate that the 30-minute limit for ULP-C was chosen for practicality reasons to expedite test administration because of the large number of people who were expected to take the test.
4. The NKSDM's documentation uses the term "sentences" without providing any further clarification; we relied on the original terminology in our review wherever possible.
5. It should be noted that the instructions for some of the sample ULP writing tasks available online "10 to 15 sentences." Although this requirement is not necessarily different from the one in NKSDM (2021e) that asks for "not fewer than 100 words," these differences in wording signify the lack of consistent instructions and complicate the appraisal of the ULP test.
6. We thank one of the anonymous reviewers for pointing this out.
7. We were able to collect some limited anecdotal evidence from several test takers who have taken the test within the past year. Some of them testified to having one short listening piece, others to having multiple pieces, yet others testified to having no listening section whatsoever.
8. The first group of raters comprised only 13 individuals on 21 July 2021.

References

Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://rm.coe.int/1680459f97

Council of Europe. (2020). *Common European Framework of Reference for languages: Learning, teaching, assessment—Companion volume with new descriptors*. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*(4), 677–714. https://doi.org/10.1017/S0140525X00043521

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *18*(2), 227–235. https://doi.org/10.1017/S1366728913000552

Kabinet Ministriv Ukrayiny. (2021a, April 14). *Postanova: Pro zatverdzhennya poryadku provedennya ispytiv na riven' volodinnya derzhavnoyu movoyu* [Decree on the procedure for conducting the exams of the level of mastery of the official language]. https://zakon.rada.gov.ua/laws/show/409-2021-%D0%BF#Text

Kabinet Ministriv Ukrayiny. (2021b, August 4). *Postanova: Deyaki pytannya provedennya ispytiv na riven' volodinnya derzhavnoyu movoyu* [Decree on conducting the exams of the level of mastery of the official language]. https://zakon.rada.gov.ua/laws/show/815-2021-%D0%BF#n12

Kunnan, A. J. (2018). *Evaluating language assessments*. Routledge. https://www.routledge.com/Evaluating-Language-Assessments/Kunnan/p/book/9780415897778

Kurteš, S., & Saville, N. (2008). The English Profile Programme—An overview. *Research Notes*, *33*, 2–4. https://www.cambridgeenglish.org/Images/23152-research-notes-33.pdf

Mazuryk, D. (2009). Kryterii vyznachennya rivnya volodinnya ukrayins'koyu movoyu yak inozemnoyu [Criteria for determining the level of Ukrainian as a foreign language]. *Theory and Practice of Teaching Ukrainian as a Foreign Language*, *4*, 27–34. http://prima.franko.lviv.ua/faculty/Philol/www/teoria_praktyka_ukr_mova/vyp_04_2009/04_mazuryk.pdf

Mazuryk, D., Antoniv, O., Synchak, O., & Boyko, H. (2018). *Standartyzovani vymogy do rivniv volodinnya ukrayins'koyu movoyu yak inozemnoyu A1-C2* [Standardized requirements for mastering Ukrainian as a foreign language A1-C2]. https://cutt.ly/tLRjSCK

Ministerstvo Osvity i Nauky Ukrayiny. (2018, May 10). *MON proponuye dlya hromads'koho obhovorennya Derzhavniy standart z ukrayins'koyi movy yak inozemnoyi* [Ministry of

Education and Science puts forward for discussion the National standard of Ukrainian as a foreign language]. https://mon.gov.ua/ua/news/mon-proponuye-do-gromadskogo-obgo-vorennya-derzhavnij-standart-z-ukrayinskoyi-movi-yak-inozemnoyi

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021a). *Vymogy do rivniv volodinnya derzhavnoyu movoyu* [Requirements for the levels of proficiency in the official language]. https://ips.ligazakon.net/document/re36547?an=1

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021b). *Poryadok perevirky rivnya volodinnya derzhavnoyu movoyu, yedyni vymogy do protsedury provedennya ispytiv ta kry-teriyi otsinyuvannya* [Procedure for determining the proficiency level in the official language, requirements for the exam administration and evaluation criteria]. https://ips.ligazakon.net/document/RE36519?an=20

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021c). *Rishennya pro zatverdzhennya klasyfikatsii rivniv volodinnya derzhavnoyu movoyu ta vymogy do nyh* [Resolution on approv-ing the classification and requirements for the levels of the mastery of the official language]. https://ips.ligazakon.net/document/re36546?an=1

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021d). *Rishennya pro zatverdzhennya Polozhennya pro ispytovi komisii* [Resolution on approving the Decree on the examination committees]. https://zakon.rada.gov.ua/laws/show/z0820-21#Text

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021e). *Prohrama ispytu na riven' volodinnya derzhavnoyu movoyu osobamy, yaki zobov'yazani volodity derzhavnoyu movoyu ta zastosovuvaty yiyi pid chas vykonannya sluzhbovyh obov'yazkiv* [Syllabus for ULP 2.0]. https://mova.gov.ua/storage/app/sites/19/exam/vikonsluzhbobovyazkiv.pdf

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2021f). *Prohrama ispytu na riven' volodinnya derzhavnoyu movoyu osobamy, yaki mayut' namir nabyty hromadyanstvo Ukrayiny* [Syllabus for ULP-C]. https://mova.gov.ua/storage/app/sites/19/exam/gromadyan-stvo.pdf

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2022a). *Harakterystyka ispytovoyi roboty ta zrazky zavdan'dlya ispytu na riven' volodinnya derzhavnoyu movoyu dlya nabuttya hromadyanstva Ukrayiny* [Sample tasks from ULP-C]. https://cutt.ly/tZri1Hg

Natsional'na Komisiya zi Standartiv Derzhavnoyi Movy. (2022b). *Zrazky zavdan'ispytu na vyznachennya rivnya volodinnya derzhavnoyu movoyu dlya vykonannya zluzhbobyh obov'yazkiv* [Sample tasks from ULP 2.0]. https://mova.gov.ua/storage/app/sites/19/exam/zavdan-ispituklyuchi-1.pdf

Ozernyi, D. M. (2021). L1 vs. L2 vs. L3 transfer: Evidence contra wholesale transfer models and privileged languages from grammatical gender and definiteness acquisition in sequen-tial quadrilinguals. *Proceedings of the Linguistic Society of America*, *6*(1), 9–23. https://doi.org/10.3765/plsa.v6i1.4891

Sherkina-Lieber, M., Perez-Leroux, A. T., & Johns, A. (2011). Grammar without speech produc-tion: The case of Labrador Inuttitut heritage receptive bilinguals. *Bilingualism: Language and Cognition*, *14*(3), 301–317. https://doi.org/10.1017/S1366728910000210

van Ek, J. A. (1975). *The Threshold Level in a European unit/credit system for modern language learning by adults*. Council for Cultural Co-operation of the Council of Europe.

Vidomosti Verhovnoyi Rady. (2019). *Zakon Ukrayiny pro zabezbechennya funktsionuvannya ukrayins'koyi movy yak derzhavnoyi* [Law of Ukraine on ensuring the use of Ukrainian as the official language]. https://zakon.rada.gov.ua/laws/show/2704-19#Text