



PRODUCTIVE MINORITY-LANGUAGE USE IN NATIONAL ASSESSMENT: IMPLICATIONS FOR BILINGUAL DEVELOPMENT

Maja Melinc Mlekuz

University of Nova Gorica, Slovenia

E-mail: maja.melinc.mlekuz@ung.si

Abstract

Assessment tasks differ in how much language students must formulate independently, and that difference shapes what an assessment can show about productive language use. This study examines the issue in minority-language assessment by analysing the complete publicly accessible corpus of Grade 8 Slovene INVALSI forms from the first computer-based cycle: two forms comprising 88 task groups. A six-level response-demand framework was developed and applied through qualitative document analysis, descriptive counts, and independent recoding. It distinguished selected and structured responses, local constructed responses, bounded sentences, relational responses, and sustained texts. Thirty-nine task groups (44.3%) required students to formulate Slovene responses. Of these, 23 required a word, phrase, grammatical form, number, or other local unit; 10 required a bounded sentence or formal transformation; and six required a causal, comparative, descriptive, or justificatory relation. No task required students to select, organise, and develop meaning across multiple sentences. The 2019 Italian INVALSI exemplar showed a similar upper boundary, whereas the 2019 Slovenian Grade 9 NPZ required four multi-sentence texts and assessed content and language quality separately. The Slovene forms sample productive language use at local, sentence, and relational levels, but not sustained academic writing. This evidential boundary matters in Slovene-medium education, where Slovene is both a minority language and the language of education. The analysis does not show that students cannot write extended texts or that unequal language exposure affected performance. Evidence of extended writing should therefore be drawn from complementary sources, including the lower-secondary State Examination and classroom writing across subjects.

Keywords: *bilingual development, document analysis, INVALSI, minority-language assessment, productive language use*

Introduction

An open question can ask for one inflected word, a sentence assembled from supplied material, or an independently organised text. Those performances do not offer the same evidence. A local answer may demonstrate lexical, semantic, or morphosyntactic control. A sustained text requires more: students must select and order content, connect propositions, maintain a topic, and manage linguistic form while meaning develops. Response demand therefore sets the limit of any inference drawn from the task (Bennett, 1993). In this article, productive use means only the language a student must supply; it is not a measure of overall proficiency.

Slovene-medium schools within Italy's public education system use Slovene as the language of education (Italian Republic, 2017), but students' exposure to Slovene and Italian varies across families, peer groups, media, and local settings (Brezigar & Verša, 2023; Grgič & Pertot, 2022; Jagodic, 2019). For some students, school is the main setting for extended academic work in Slovene. These studies do not report performance on the forms analysed here and cannot show that unequal exposure affected results. Their role is contextual: they explain

why it matters whether students are asked to explain, compare, justify, and develop subject knowledge in Slovene. Functional bilingual development is used here in a limited, educational sense: both languages remain available for the activities required in school and social life. It does not mean identical attainment in Italian and Slovene or conformity to two monolingual standards (Grgič & Pertot, 2022). The documents cannot establish whether individual students are balanced bilinguals. They can show whether the assessment calls on Slovene for particular academic practices, including the independent development of meaning across a text.

Validity claims must match the performance elicited by a task and the features recognised in scoring (American Educational Research Association et al., 2014; Kane, 2013; Messick, 1995). A selected-response item may demand complex inference, whereas an open item may ask for one word recoverable from the source. Language demands can affect constructed-response performance even when language is not the intended construct (Noble et al., 2023), and INVALSI research has found considerable variation in semantically narrow open answers (Leonetti & Viale, 2020). Studies of orthographic items likewise show that content, format, correction criteria, and processing demands must be considered together (Lo Duca & Toth, 2021). The labels open and closed are too crude for this purpose; the relevant question is what language the student must produce.

Problem Statement

Counting open questions does not reveal how much independent language a test requires. In the 39 task groups with student-supplied Slovene, the minimum fully creditable answer ranged from one lexical or grammatical unit to a sentence or a brief relation among ideas. Treating these responses as equivalent would blur the difference between local production and independent writing. In a minority-language school system, that difference affects the claims that can be made from assessment results: a word, a transformed sentence, or a short justification can demonstrate useful productive control without showing that a student can select, organise, and sustain meaning across a text. The study therefore identifies the highest productive demand in the two publicly released forms and asks what that limit permits us to say about language competence, functional bilingual development, and the need for complementary assessment.

Assessment Context and Research Gap

INVALSI was not designed to test composition. Its Grade 8 Italian framework covers reading comprehension, vocabulary, and reflection on language; extended writing lies outside the stated construct (INVALSI, 2018a). That boundary reflects the purpose and practical constraints of large-scale assessment, not a judgement that writing is unimportant. In the survey reported by Cardone et al. (2017), most lower-secondary teachers considered the reading-comprehension and grammar items consistent with work carried out during the school year. Computerisation in 2018 changed how open responses could be processed (INVALSI, 2018b). Marsili et al. (2020) report that centralised, supervised processing reduced the person-hours required once manual coding became difficult to sustain. Leonetti and Viale (2020) found substantial linguistic variation even in semantically narrow Grade 8 answers, while Le Rose and Ricciardi (2021) distinguish structured testing from written composition. Together, these studies show why response format must be interpreted through the evidence it can provide.

At the same stage of schooling, students take the lower-secondary State Examination at the end of the first cycle of education (*Esame di Stato conclusivo del primo ciclo di istruzione*). The examination comprises three written papers and an oral interview. The first written paper is in Italian or the language of education, so students in Slovene-medium schools write in Slovene. School examination boards prepare prompts within nationally defined categories: narrative or

descriptive writing, argumentation, and the comprehension and synthesis of literary, expository, or scientific texts; combined tasks are also permitted (Italian Ministry of Education, 2017, 2018). Unlike INVALSI, the first written paper requires extended, independently organised writing. The absence of a multi-sentence task from the released INVALSI forms therefore does not mean that sustained writing is absent from end-of-cycle assessment or Slovene-medium education.

The move to TAO made the difference between an open answer and independent writing especially visible in the Slovene forms. Four forms were prepared in 2018/2019: two were entered in that cycle, and two were adapted to INVALSI's task typology and technical specifications before entry in 2019/2020. The 2020 administration was cancelled during the COVID-19 pandemic; testing resumed in 2021 (INVALSIopen, 2021). Before computerisation, Slovene tests contained many open questions, but these often required copying from a source or producing a longer yet locally bounded answer rather than organising a text independently. During the transition, the team was encouraged to use closed or tightly constrained formats where possible so that responses could be processed consistently. This procedural account comes from the researcher's contemporaneous professional knowledge. It explains how the research question arose but is not evidence in the document analysis.

Research on the Slovene version has addressed the institutional organisation and general task profile of the assessment (Kaluža Pocecco, 2016), mathematics items in Slovene-medium schools (Doz, 2019), and the use of INVALSI results to strengthen the school's language of education (Kovic & Cividin, 2022). No study has yet mapped the response demand of every publicly accessible Slovene Grade 8 task group or located the point at which local production gives way to sustained text construction. The present analysis addresses that gap with a response-demand framework applied to the full accessible corpus. It does not assess the overall cognitive difficulty or psychometric quality of the items.

The 2019 Slovenian Grade 9 National Assessment of Knowledge (NPZ) provides a deliberately limited contrast because it belongs to another national system and follows a different design. Analyses of Slovene NPZ tasks have identified independent writing and text-based argumentation as recurring areas of difficulty (Cestnik et al., 2012). Rosc-Leskovec (2013) discussed students' language ability through NPZ achievement data, while Smolej (2021) reported persistent problems with coherent, linguistically accurate texts, summarising, drawing conclusions, and using source information. Those studies concern student performance. In the present analysis, the NPZ prompts and scoring instructions are used only to establish what the 2019 instrument explicitly required and rewarded at text level. The two Slovene INVALSI forms remain the primary corpus and were coded in full; the Italian exemplar and the Slovenian NPZ serve only as released 2019 comparison sets. They are not treated as psychometrically equivalent, and the findings are not generalised to later INVALSI cycles, the full Italian item bank, or the Slovenian assessment system as a whole.

Aim and Research Questions

The study pursued three aims: to identify the highest level of independent Slovene formulation required in the two publicly released Grade 8 INVALSI forms; to examine the tasks nearest the boundary between a local answer and a text; and to compare that boundary with two official document sets from 2019.

RQ1. What levels of independent Slovene-language formulation were required in the two analysed Grade 8 forms?

RQ2. Where did productive demand stop: at a local word or phrase, a bounded sentence or formal transformation, a local relation among content elements, or an independently organised text?

RQ3. How did that upper boundary compare with the released 2019 Italian Grade 8 INVALSI exemplar and the 2019 Slovenian Grade 9 NPZ?

Research Methodology

Design and Documentary Logic

The unit of interest was the linguistic work needed for a fully creditable answer, not the topic of a task or its nominal format. To capture that work, the study used qualitative document analysis supported by descriptive counts (Bowen, 2009). The documents were retrieved, coded, compared, and audited between June and August 2026. The numbered task group served as the unit of analysis because it was the only stable, visible unit shared by both Slovene forms and their answer documentation. The analysis concerns the evidence that tasks can elicit; it does not address student achievement, item difficulty, psychometric properties, or teaching effects.

Each task group was classified through three questions: Was the response selected from supplied material or newly worded? If wording was required, how much content and sentence structure had the task already determined? Did the student have to select and order information beyond the immediate answer? The assigned level reflected the minimum response eligible for full credit under the prompt and scoring materials. Neither answer-box size nor the platform's technical label counted as evidence. The cross-document comparison examined whether wording was supplied, whether a complete clause or sentence was required, whether content had to be organised across sentences, and whether content, organisation, and language quality were scored separately.

Documents and Analytical Boundaries

The primary corpus comprised two Slovene-language Grade 8 INVALSI forms from the 2018/2019 assessment cycle and their accompanying answer documentation, downloaded from the official INVALSI materials on 15 June 2026 (INVALSI, 2019a, 2019b). Each form contained 44 numbered task groups, giving 88 groups in total. The tasks were organised around six sets of source texts and two grammar-and-language sections, and every group was included. These were the complete Slovene Grade 8 forms that could be identified as publicly available from the computer-based period. Access and test-security constraints determined their inclusion; the observed response profile did not. Both forms were entered during the Slovene team's early work with TAO, when INVALSI staff in Rome provided direct technical support, and were later released as examples. Subsequent Slovene forms remain in the confidential INVALSI item bank and may be reused. They cannot therefore be reproduced, described task by task, or added to the corpus. The analysis covers the full accessible set but cannot establish whether later confidential forms follow the same pattern.

Two further sets of official 2019 documents were used for comparison. The Italian set consisted of the correction grid for the published Grade 8 INVALSI exemplar, which records the format and expected answer for 40 questions (INVALSI, 2019c). The Slovenian set comprised the Grade 9 NPZ Slovene test and its moderated scoring instructions (Državni izpitni center, 2019a, 2019b). Although the grade labels differ, both assessments occur at the end of lower-secondary or basic schooling in their respective systems. The shared year and broadly comparable curricular position allow a limited documentary comparison, but not psychometric equivalence or direct comparison of score scales. The comparison served only to identify the highest response demand visible in each set of released materials.

Table 1
Documentary Design and Permitted Comparisons

Document set	Role in the study	Analysis performed	Permitted claim
Two publicly released Slovene Grade 8 INVALSI forms (88 task groups)	Primary corpus: the full accessible Slovene Grade 8 CBT set	All 88 groups coded on six levels; 18 borderline cases examined in detail	Claims are limited to the response demands in these two public forms
2019 Italian Grade 8 INVALSI exemplar (40 questions)	Comparison within the same assessment system and year	Official format labels and expected answers counted and examined	Shows the highest response demand in this exemplar only
2019 Slovenian Grade 9 NPZ Slovene test and scoring instructions	Comparison in the same language at the end of basic schooling	Text-production prompts and scoring criteria examined	Provides an example of direct assessment of content organisation and language quality

Note. The comparison concerns released documents and response demands. It does not equate the purposes, populations, score scales, or psychometric designs of the three instruments.

Response-Demand Framework

The framework places responses on six levels according to how much wording, content, and organisation the student must supply. A selected-response item may be cognitively demanding, yet it remains at Level 0 or 1 when the answer wording is supplied. Conversely, a grammatically complete sentence does not reach Level 4 or 5 simply because it is long; the task may already have fixed its content and structure. The levels are not CEFR proficiency levels. The framework draws instead on a broader progression also visible in the CEFR writing scales, from isolated phrases and simple sentences to connected, structured, and purposefully developed texts (Council of Europe, 2020; North & Schneider, 1998). In an analysis of 1,155 argumentative essays by adolescent EFL learners, Khushik and Huhta (2020) found that CEFR levels A1, A2, and B1 were distinguished most clearly by features of the produced discourse, including the length of production units, subordination, and phrasal density. This research supports the need for actual text production when the intended inference concerns extended writing. It is not used here to assign CEFR levels to INVALSI tasks or to students.

Table 2
Six Levels of Response Demand

Level	Required response	Evidence the task can elicit	Boundary
0. Selected	Select one supplied option	Recognition or discrimination	No student-supplied wording
1. Structured	Match, order, classify, or manipulate supplied elements	Structured processing of supplied material	Wording remains supplied
2. Local constructed	Supply a word, form, number, short phrase, or other local unit	Local lexical, semantic, numerical, or formal production	No necessary complete-sentence organisation
3. Sentence/formal	Produce a bounded clause or sentence, or transform a supplied sentence	Sentence-level formulation or formal control	Content remains local or substantially supplied
4. Relational	Connect several content elements or state a cause, contrast, comparison, or judgement with a reason	Local integration and explicit relation	No sustained organisation beyond the immediate answer
5. Sustained text	Select and organise content in a coherent multi-sentence text	Text-level organisation, progression, and extended writing	Requires organisation and development across multiple sentences; scoring may distinguish content, organisation, and language quality

Note. The levels describe the independence and extent of the required response, not overall cognitive difficulty.

Coding, Review, and Comparison Procedure

Coding began with an inventory of every task group by form, section, task number, source text, interaction type, and documented answer. Each group was assigned the highest level required by any subpart, based on the minimum response eligible for full credit. All 39 constructed-response groups were checked again against the original prompt and answer documentation. The 18 groups labelled as extended constructed responses in the preliminary five-level scheme were retained for the boundary audit. For each, the record included the productive function, minimum semantic components, use of source information, principal operation, degree of constraint, scoring evidence, and evidential limit. These cases were not selected after the distribution was known. The final totals were reconciled with 44 groups in each form and 88 overall.

Before reliability was calculated, a subject specialist reviewed the full coding table with the researcher to refine the provisional framework. Eight borderline cases were examined in detail. Four provisional codes changed, and the distinction between adjacent levels was made more explicit. The audit trail preserves the original, alternative, and final decisions. Because the specialist had seen the researcher's coding and the decisions were reached through discussion, this stage was treated as framework development rather than independent coding and was excluded from the reliability calculations.

After the framework had been finalised, a second coder independently classified all 88 task groups in random order. The coder had no access to the researcher's classifications or aggregate findings, and no individual task was discussed in advance. Before adjudication, the coders agreed exactly on 72 of the 88 groups (81.8%) and differed by no more than one level on 84 (95.5%). Unweighted Cohen's κ was .746 and linearly weighted κ was .834. Twelve of the 16 disagreements involved adjacent levels, including five at the boundary between Levels 3 and 4; four differed by more than one level. Following discussion of each case against the fixed criteria, the two coders reached a final classification. In all 16 cases, the adjudicated classification coincided with the researcher's original code. The reported distributions and conclusions were therefore unchanged, and the full adjudication record was retained.

For the Italian comparison, the 40 entries in the official correction grid were counted by item format, and their expected answers were examined against the same four documentary dimensions. For the Slovenian NPZ, the analysis identified prompts that explicitly required a coherent or short text and examined how the scoring instructions treated content, text organisation, and language quality. Point values were taken directly from the official scoring document. No inferential statistics were used.

Researcher Positionality and Ethics

The researcher's involvement in the Slovene INVALSI process began in 2019, while she was working at the Slovene Research Institute (SLORI) in Trieste. Her responsibilities included coordinating the teachers who developed the Slovene-language assessment tasks, communication with the Regional School Office in Trieste and INVALSI in Rome, and entering the assessment materials into the TAO platform. The Slovene team prepared four forms: two were entered in the 2018/2019 cycle, and two were adapted to INVALSI's task typology and technical specifications before entry in 2019/2020. The analysis reported here uses no internal development records, student results, or unpublished task materials. Its corpus is limited to the two forms and answer documentation later released to the public. In subsequent cycles through 2022, the researcher's role expanded to include substantial involvement in writing, editing, selecting, and assembling tasks, alongside coordination and TAO entry. Those materials remain confidential and are not analysed. This experience informed the research question but also introduced a possible risk of confirmation bias. Explicit coding rules, a preserved audit trail, specialist review during framework development, and blind independent coding of the complete public corpus were used to reduce that risk.

The study involved no human participants, interventions, personal data, or identifiable information about students or schools; ethical approval and informed consent were therefore not required. OpenAI's ChatGPT was used for translation into English and language editing.

Research Results

Response-Demand Profile in the Slovene Forms

Neither form required a sustained text. Each contained 44 task groups. Form 1 comprised 19 selected-response groups, seven structured or supplied-response groups, 10 local constructed responses, five bounded sentence or formal responses, and three relational responses. Form 2 contained 18, five, 13, five, and three groups at the respective levels. Across both forms, 49 of the 88 task groups (55.7%) supplied the response material, while 39 (44.3%) required students to produce language of their own. Among those 39 groups, 23 ended with a word, phrase, number, grammatical form, or another local unit; 10 required a bounded sentence or the transformation of a complete sentence; and six required students to express a relation between content elements. None required several sentences to be organised into a text.

Table 3*Distribution of Response Demands in the Two Slovene Forms*

Level	Form 1, n (%)	Form 2, n (%)	Total, n (%)
L0 selected	19 (43.2)	18 (40.9)	37 (42.0)
L1 structured	7 (15.9)	5 (11.4)	12 (13.6)
L2 local	10 (22.7)	13 (29.5)	23 (26.1)
L3 sentence/formal	5 (11.4)	5 (11.4)	10 (11.4)
L4 relational	3 (6.8)	3 (6.8)	6 (6.8)
L5 sustained text	0	0	0
Total	44 (100)	44 (100)	88 (100)

Note. Each numbered group was classified by the highest response demand represented. Levels 0–1 total 49 groups (55.7%); Levels 2–4 total 39 groups (44.3%). Percentages are within form or within the combined corpus.

A Complete Sentence Did Not Necessarily Mean Independent Content

Level 2 included genre labels, dates, percentages, pairs of synonyms, and grammatical forms inserted into sentences already provided. Students had to retrieve Slovene words or supply the required inflection, but the productive work remained local. A group with several gaps was not classified as text construction when each gap required a separate word or form.

Six of the 10 Level 3 groups required a complete sentence to be transformed. Across the two forms, students changed direct speech into reported speech, punctuated a supplied sentence, or changed a sentence into the dual. The resulting sentence could be grammatically complete, but its proposition, vocabulary, and sequence had already been provided. These tasks showed control of reported speech, punctuation, or number agreement rather than independent development of content. The other four Level 3 groups required a bounded semantic response. Two asked for a sentence that paraphrased a local proposition or expression. One required a full statement of the evidence that Joshua Bell was a leading violinist; another required a sentence linking the title ‘Seven Minutes of Terror’ to the dangerous landing phase on Mars. These tasks allowed more freedom of wording than the formal transformations, but each remained confined to one local proposition.

The Six Relational Responses

Six groups reached Level 4, three in each form. In Form 1, students explained why infected corpses catapulted over the walls of Caffa could be regarded as biological warfare, decided whether Giovanni Villani had completed his chronicle and justified their answer, and linked acid rain, salt-bearing rocks, and sea salinity. In Form 2, they combined details to describe Ema’s appearance, brought together several physical features of the creatures pursuing her, and identified the relevant difference between Ema’s and Peter’s dreams. What distinguished these tasks was the need to state an explicit relation: a cause or mechanism, a claim supported by a reason, a comparison, or an integrated description. Even so, each response remained tied to one immediate question. None required the explanation or comparison to be developed through a sequence of independently organised sentences.

Table 4
The 18 Boundary Cases: Exact Productive Demand

Form/task (level)	Required action	Minimum evidence for full credit	Why not sustained text
F1 A4 (L3)	Restate one local proposition	Mother taught the kittens to flee from the writer	One proposition; no development
F1 A6 (L3)	Paraphrase an expression	Explain what it meant to ignore the cat	One local semantic reformulation
F1 B5 (L4)	Explain a mechanism	Link infected corpses, deliberate projection, and disease transmission	Local causal explanation
F1 B9 (L4)	Judge and justify	No, because Villani died first	One claim–reason pair
F1 C6 (L4)	State a causal relation	Connect acid rain, salt-bearing rocks, and sea salinity	Local causal chain
F1 D5 (L3)	Transform reported speech	Rewrite one supplied sentence	Content and sequence supplied
F1 D6 (L3)	Insert punctuation	Rewrite one supplied sentence with commas	No content selection
F1 D7 (L3)	Change grammatical number	Rewrite one supplied sentence in the dual	No content selection
F2 A3 (L4)	Integrate description	Combine hair, dark circles, and fear in the eyes	Several details, but one local description
F2 A7 (L4)	Integrate description	Combine several physical features of the creatures	Several details, but one local description
F2 A9 (L4)	Compare and justify	State the difference between Ema's and Peter's dreams	One comparison
F2 A11 (L2)	Supply two descriptors	Dirty and muddy	Short local description
F2 B8 (L3)	State relevant evidence	The award as evidence of Bell's standing	One evidence sentence
F2 B10 (L2)	Reformulate one consequence	The artworks lose their value	One local proposition
F2 C1 (L3)	Link title and referent	Connect the title to the seven-minute landing	One sentence-level link
F2 D5 (L3)	Transform reported speech	Rewrite one supplied sentence	Content and sequence supplied
F2 D6 (L3)	Insert punctuation	Rewrite one supplied sentence with commas	No content selection
F2 D7 (L3)	Change grammatical number	Rewrite one supplied sentence in the dual	No content selection

Note. The table includes all 18 groups retained for detailed boundary analysis. English descriptions paraphrase the Slovene prompts and scoring evidence.

Comparison with the 2019 Italian INVALSI Exemplar

The official correction grid for the 2019 Italian Grade 8 exemplar contained 40 questions, 11 of which were labelled *aperta univoca* [open-univocal]. Ten expected a word, a short phrase, or a pair of adjectives; examples included ‘fascismo’, ‘risorse idriche’, ‘universo’, ‘matematica’, and the pair ‘polemico’ and ‘vendicativo’. The remaining question required one sentence explaining that italics marked a quotation from Galileo’s *Saggiatore*. None required a multi-sentence text (INVALSI, 2019c). The Italian exemplar places the same upper boundary within the wider INVALSI design. Its Grade 8 framework centres on reading, vocabulary, and reflection on language. In the published exemplar, productive demand ended with local wording and one bounded explanation. The comparison cannot establish that either test is better or more representative. It does, however, narrow the permissible inference from the Slovene forms: the released tasks do not sample students’ ability to organise extended subject knowledge in Slovene.

Comparison with the 2019 Slovenian NPZ

The 2019 Slovenian Grade 9 NPZ followed a different design and included four explicit writing tasks. In the literary section, students produced a coherent analysis of a poem, addressing the speaker’s changed world, the role of nature, and two poetic devices supported by examples. They then compared two poems in a coherent text, explaining how love affected both speakers, comparing their modes of expression, and justifying which poem they found more persuasive. In the non-literary section, students wrote a short explanatory text about language choice and action despite dyslexia, followed by a coherent profile of Gojmir Lešnjak Gojc in no more than six sentences and based on four specified content areas (Državni izpitni center, 2019a).

The scoring instructions assessed dimensions absent from the INVALSI materials. The four tasks accounted for 20 of the available 55 points (36.4%): 12 for the required content and eight for grammatical and orthographic accuracy and stylistic appropriateness. Guidance for markers addressed word order, repeated words and sentence patterns, overuse of conjunctions, irrelevant digressions, excessive length, and the difference between a coherent text and a list of bullet points (Državni izpitni center, 2019b). The Slovene and Italian INVALSI materials did not require students to demonstrate these aspects of writing. The mode of administration also differed. The 2019 Slovenian NPZ was a paper-and-pencil test, whereas the INVALSI forms were delivered online. For this comparison, two design features matter most: the NPZ required four multi-sentence texts and awarded separate points for content and language quality. Students had to select and connect information across sentences, and the scoring scheme assessed both what they wrote and how they expressed it.

Table 5
Upper Boundary of Productive Demand in the Three Released 2019 Document Sets

Document set	Productive responses	Highest demand	Text production and scoring
Two Slovene Grade 8 INVALSI forms	39/88 groups: 23 local, 10 sentence/formal, 6 relational	Relational response (Level 4)	No; not applicable because no text was required
Italian Grade 8 INVALSI published exemplar	11/40 open-univocal: 10 word/phrase, 1 sentence	One bounded explanatory sentence	No; answer key only, with no text-level criteria
Slovenian Grade 9 NPZ Slovene test	Local and sentence responses present; not exhaustively coded	Sustained text production	Yes: 4 tasks, 20/55 points; content and language quality scored separately; 8 language points

Note. The table compares documented features rather than rates. The denominators, purposes, score scales, and psychometric designs differ across the three document sets. Italian counts refer to one published exemplar; NPZ figures refer to four explicitly identified text-production tasks.

Discussion

What the Slovene Forms Assess

Calling all 39 groups constructed responses conceals the central result: their productive demands are not equivalent. They range from genre labels, dates, percentages, synonyms, and inflected forms to local paraphrases, transformed sentences, and six causal, comparative, descriptive, or justificatory relations. Taken together, the tasks sample local lexical, semantic, morphosyntactic, and relational production. They do not require independent text construction. In the six sentence-transformation tasks, both the proposition and the lexical sequence were supplied. In the six Level 4 groups, students connected information only within the immediate answer. No task required them to select and order several ideas, maintain a topic, or control language across a multi-sentence whole. A multiple-choice item may require complex inference, while a short open answer may be straightforward. Interpretation should therefore follow the language performance elicited and recognised in scoring (Leonetti & Viale, 2020; Lo Duca & Toth, 2021).

Selected and tightly constrained responses are useful components of large-scale assessment. Such instruments must balance construct coverage against comparability, scoring consistency, time, and cost. Marsili et al. (2020) show how these pressures shaped the processing of open INVALSI responses after computerisation. In another field, Slepko and Shiell's (2014) comparison of response formats illustrates the additional demands that constructed responses place on time, cost, and inter-rater reliability. Their study does not concern INVALSI, but it clarifies the general design trade-off. These design choices set the scope of valid claims. Reading comprehension, reflection on language, and brief productive responses remain valuable parts of the assessment.

The Text-Level Gap and Bilingual Development

The Italian exemplar places the Slovene finding within the wider INVALSI design rather than marking it as a peculiarity of the minority-language forms. Both document sets reflect a construct centred on reading, vocabulary, and reflection on language rather than composition. In Slovene-medium education, however, that boundary has a specific consequence. Italian is strongly supported in the wider environment, while opportunities to use Slovene outside school vary across families, peer groups, media, and local communities. The contextual studies do not show that these differences affected performance on the analysed forms; they explain why evidence about sustained writing in Slovene cannot be assumed. Sustained writing is not the whole of communicative competence, but it is one form of higher written productive ability. In this article, functional bilingual development means that Slovene remains available for extended explanation, interpretation, argument, and the organisation of knowledge, not that students attain symmetrical proficiency in every domain. A genre label, transformed sentence, or short justification may show useful local control. It cannot show whether a student can select evidence, maintain a line of reasoning, connect propositions, and manage language as a text develops. The two forms leave that academic use of Slovene unexamined.

The boundary identified here concerns the INVALSI evidence. Extended writing remains part of Slovene-medium education through the written Slovene paper in the lower-secondary State Examination, which requires extended production within nationally defined text types (Italian Ministry of Education, 2017, 2018). The instruments serve different purposes: INVALSI samples reading, reflection on language, information processing, and brief productive responses, whereas the lower-secondary State Examination and classroom assessment can show how students organise a text independently. A comprehensive interpretation of written Slovene competence must use these sources together.

Assessment and Teaching Implications

Assessment reports should describe INVALSI results within the boundaries of the stated construct and should not present them as a complete measure of written Slovene competence. Decisions that depend on evidence of sustained writing must also use the written Slovene paper in the lower-secondary State Examination and classroom assessment. Where a common supplementary measure is needed, a separately scored, source-based writing task could distinguish required content, organisation and coherence, and language accuracy. In the context of Slovene as a second language, Ferbežar (2024) similarly shows why successful task completion cannot be reduced to linguistic correctness. The proposal here is not to turn INVALSI into a composition examination, but to coordinate evidence from instruments that serve different purposes.

The existing constructed-response tasks can serve as starting points for longer classroom writing. Students might first identify one relevant detail, then state a relation between two details, and finally combine several relations in a paragraph or short text. The Level 4 tasks offer ready examples: the causal link between acid rain and salinity, the judgement about Villani and its justification, and the comparison of two dreams could each be developed from a local answer into a fuller explanation. More open questions alone will not create this progression; teaching must move deliberately from local formulation to connected discourse. This progression should not be confined to language lessons. Slovene is also the language through which students learn and demonstrate subject knowledge. Qi and Fang's (2023) qualitative study involved English learners in two science classrooms in the United States and does not provide direct evidence about Slovene-medium schools. Its distinction between writing of at least a paragraph and restricted activities involving little composition is nonetheless useful: extended writing

allows students to organise, connect, and develop disciplinary meaning rather than display isolated pieces of knowledge. Slovene-medium schools therefore need sustained explanation, interpretation, and argument in language and content subjects, not only during preparation for the final examination.

Limitations and Contribution

The corpus is necessarily small: two Slovene forms from one cycle. They are, however, the complete Grade 8 Slovene forms publicly available for task-level analysis from the first computer-based cycle; later forms remain in a confidential item bank and may be reused. The Italian comparison draws on one published exemplar rather than the full operational bank, and the NPZ differs in country, grade, curriculum, purpose, administration, and scoring. These documents reveal contrasting designs but cannot show how common either design is across the two systems or support a ranking of the tests. The account of the lower-secondary State Examination is based on the national regulatory framework, not a corpus of school-prepared Slovene papers. It establishes what the examination may assess, not how often particular task types occur or how well they are implemented.

Because no student responses or scores were analysed, the study cannot establish item difficulty, the length of answers actually written, scoring reliability in operational use, subgroup differences, or whether students produced more than the prompts required. Agreement between the two coders was high overall, although five disagreements at the boundary between Levels 3 and 4 show that bounded formulation and local integration are the hardest categories to separate. Table 4 makes those judgements available for inspection. The complete anonymised coding and reliability record has been retained and can be supplied on request. Within these boundaries, the complete task-group analysis makes a specific contribution. It distinguishes in a replicable way among local answers, bounded sentences, relational responses, and independently organised texts in the publicly accessible Slovene Grade 8 INVALSI corpus. The Italian exemplar locates the same ceiling in released INVALSI material, while the Slovenian NPZ and the lower-secondary State Examination illustrate two ways of requiring and assessing text-level content decisions and language quality. The findings concern the evidence produced by different assessment designs; they do not make claims about student achievement.

Conclusions

Across the complete publicly accessible Slovene Grade 8 INVALSI corpus, productive demand stopped at the level of a local relation. Of 88 task groups, 39 required students to supply Slovene: 23 ended with a word, phrase, number, grammatical form, or another local unit; 10 required a bounded sentence or formal transformation; and six required a causal, comparative, descriptive, or justificatory relation. None required a multi-sentence text. The published Italian Grade 8 exemplar reached a similar limit, whereas the 2019 Slovenian Grade 9 NPZ required four texts and assessed content and language quality separately. The Slovene INVALSI forms therefore sample productive language use at local, sentence, and relational levels, but not sustained academic writing.

This ceiling describes the evidential reach of the released forms. It should not be interpreted as a claim about students' ability to write. The stated INVALSI construct remains intact. The released forms nevertheless leave one academic use of Slovene outside the evidence: selecting and organising ideas across a sustained text. In a minority-language setting, that boundary matters because Slovene must remain available for organising knowledge, sustaining explanations, and developing arguments across subjects.

INVALSI results should therefore be read alongside the written Slovene paper in the lower-secondary State Examination and classroom writing. Teachers can use local answers as starting points for paragraphs and longer texts in language lessons and content subjects. Policymakers and assessment reports should not present INVALSI as a comprehensive measure of written Slovene competence.

Author Notes

Acknowledgements. The author is grateful to the Slovene Research Institute (SLORI) in Trieste for its support and especially to her colleague Martina Pettiroso for her collaboration in the INVALSI-related work.

Funding. The research work undertaken in preparation for this article at the University of Nova Gorica was partly supported by the Slovenian Research and Innovation Agency (ARIS) through the research programme Theoretical and Experimental Linguistics (P6-0382).

Ethics Statement. The study used publicly released or publicly documented assessment materials. It involved no human participants, personal data, student responses, or identifiable school results; institutional ethical approval and informed consent were therefore not required.

Data Availability. All source documents are identified in the References. The complete coding matrix, confusion matrix, reliability calculations, and adjudication record can be supplied in anonymised form on request. The workbook is not publicly deposited because it contains task-level wording from copyrighted assessment materials.

Conflict of Interest and Positionality. The Methodology describes the researcher's professional involvement in the Slovene INVALSI process. The analysis uses only publicly released documents; no internal development records, student data, school results, or confidential task materials were included.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Bennett, R. E. (1993). On the meanings of constructed response. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1–27). Lawrence Erlbaum Associates.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Brezigar, S., & Verša, A. (2023). Slovensko šolstvo v Italiji: Včeraj, danes in jutri [The Slovene school system in Italy: Yesterday, today, and tomorrow]. In A. Istenič, M. Gačnik, B. Horvat, M. Kukanja Gabrijelčič, V. R. Kiswarday, M. Lebeničnik, M. Mezgec, & M. Volk (Eds.), *Vzgoja in izobraževanje med preteklostjo in prihodnostjo* (pp. 311–327). Založba Univerze na Primorskem. <https://doi.org/10.26493/978-961-293-255-8.311-327>
- Cardone, M., Le Rose, G., & Ricciardi, V. (2017). Attività didattica e prove INVALSI: Alcune riflessioni a partire dall'opinione degli insegnanti [Teaching activity and INVALSI tests: Reflections based on teachers' views]. *Pedagogia più Didattica*, 3, 1–18.
- Cestnik, M., Medved-Udovič, V., & Žagar Pernar, T. (2012). Nacionalno preverjanje znanja iz slovenščine ob koncu tretjega obdobja osnovne šole – analiza [National assessment of Slovene at the end of the third cycle of primary school: An analysis]. *Jezik in slovstvo*, 57(1–2), 97–109. <https://doi.org/10.4312/jis.57.1-2.97-109>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Doz, D. (2019). The national assessment of mathematics in high schools in Italy with Slovene as the language of instruction. *Revija za elementarno izobraževanje*, 12(2), 155–176. <https://doi.org/10.18690/rei.12.2.155-176.2019>

- Državni izpitni center. (2019a). *Slovenščina: Preizkus znanja, nacionalno preverjanje znanja v 9. razredu* [Slovene: National assessment test, Grade 9]. <https://www.ric.si/mma/n191-101-3-1/2019061310501326/?m=1560415813>
- Državni izpitni center. (2019b). *Slovenščina: Navodila za vrednotenje, nacionalno preverjanje znanja v 9. razredu* [Slovene: Scoring instructions, Grade 9 national assessment]. <https://www.ric.si/mma/n191-101-3-2/2019061310501490/?m=1560415814>
- Ferbežar, I. (2024). (O)pravilnost na izpiti iz znanja slovenščine kot drugega jezika [(In)correctness in examinations of Slovene as a second language]. In S. Štumberger (Ed.), *Predpis in norma v jeziku* (Obdobja 43, pp. 85–92). Založba Univerze v Ljubljani. <https://doi.org/10.4312/Obdobja.43.85-92>
- Grgič, M., & Pertot, S. (2022). *L'adolescenza bilingue = Dvojezično najstništvo* [Bilingual adolescence]. Slovenski raziskovalni inštitut / Istituto sloveno di ricerche.
- INVALSI. (2018a). *Quadro di riferimento delle prove INVALSI di italiano* [Reference framework for the INVALSI Italian-language tests]. https://www.invalsi.it/wp-content/uploads/2024/09/QdR_ITALIANO.pdf
- INVALSI. (2018b). *Rapporto prove INVALSI 2018* [Report of the 2018 INVALSI assessments]. <https://www.invalsiopen.it/risultati/archivio-risultati-invalsi/rapporto-invalsi/>
- INVALSI. (2019a). *[Grade 8 Slovene-language assessment form 1 and corresponding answer documentation]* [PDF]. Retrieved June 15, 2026, from <https://www.invalsi.it/rilevazioni-nazionali/prove-invalsi-esempi/>
- INVALSI. (2019b). *[Grade 8 Slovene-language assessment form 2 and corresponding answer documentation]* [PDF]. Retrieved June 15, 2026, from <https://www.invalsi.it/rilevazioni-nazionali/prove-invalsi-esempi/>
- INVALSI. (2019c). *Griglia di correzione – Esempio di prova INVALSI di Italiano per la classe terza secondaria di primo grado* [Correction grid: Italian INVALSI exemplar for Grade 8]. https://www.invalsiopen.it/prove_statistiche/G08/ITA/2019/Griglia_correzione_ITA_08_EX1.pdf
- INVALSIopen. (2021). *I risultati delle Prove INVALSI 2021* [Results of the 2021 INVALSI assessments]. <https://www.invalsiopen.it/risultati/risultati-prove-invalsi-2021/>
- Italian Ministry of Education. (2017). *Decreto ministeriale 3 ottobre 2017, n. 741: Esame di Stato conclusivo del primo ciclo di istruzione* [Ministerial Decree No. 741 of 3 October 2017: Lower-secondary State Examination at the end of the first cycle of education]. <https://www.mim.gov.it/-/d-m-741-del-3-10-2017-esame-di-stato-conclusivo-del-primo-ciclo-di-istruzione>
- Italian Ministry of Education. (2018). *Documento di orientamento per la redazione delle tracce della prova scritta di italiano per l'esame di Stato conclusivo del primo ciclo di istruzione* [Guidance for preparing the written language paper for the lower-secondary State Examination at the end of the first cycle of education] (Nota prot. n. 892, 17 January 2018). <https://www.mim.gov.it/-/prova-scritta-di-italiano-esame-di-stato-primo-ciclo-di-istruzione>
- Italian Republic. (2017). Decreto legislativo 13 aprile 2017, n. 62: Norme in materia di valutazione e certificazione delle competenze nel primo ciclo ed esami di Stato (Art. 24). *Gazzetta Ufficiale della Repubblica Italiana*, Serie Generale, 112. https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.codiceRedazionale=17G00070&art.dataPubblicazioneGazzetta=2017-05-16&art.idArticolo=24
- Jagodic, D. (2019). Znanje in raba slovenskega jezika med mladimi v slovenskem zamejstvu v Italiji [Knowledge and use of Slovene among young people in the Slovene minority in Italy]. In S. Novak Lukanović (Ed.), *Jezikovni profil mladih v slovenskem zamejstvu* (pp. 67–120). Inštitut za narodnostna vprašanja; Slovenski znanstveni inštitut; Slovenski narodopisni inštitut Urban Jarnik; Slovenski raziskovalni inštitut.
- Kaluža Pocecco, T. (2016). Nacionalni preizkus znanja iz slovenščine ob zaključku srednje šole prve stopnje s slovenskim učnim jezikom v Italiji [The national examination of Slovene at the end of lower-secondary school with Slovene as the language of instruction in Italy]. *Revija za elementarno izobraževanje*, 9(4), 105–122.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Khushik, G. A., & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference levels A1, A2, and B1. *Applied Linguistics*, 41(4), 506–532. <https://doi.org/10.1093/applin/amy064>
- Kovic, E., & Cividin, A. (2022). I dati INVALSI come leva nella trasformazione dell'approccio al potenziamento della lingua veicolare della scuola? Evidenze dalla scuola italiana di lingua slovena nella regione Friuli-Venezia Giulia [Can INVALSI data drive changes in strengthening the school's language of education? Evidence from Italian schools with Slovene as the language of education in Friuli Venezia Giulia]. In P. Falzetti (Ed.), *Le buone pratiche all'interno della scuola* (pp. 41–58). FrancoAngeli.

- Le Rose, G., & Ricciardi, V. (2021). Prova strutturata vs prova con svolgimento aperto: Gli effetti sul rendimento degli studenti in Italiano [Structured task versus open-ended task: Effects on students' achievement in Italian]. *Italiano a scuola*, 3(1), 83–98. <https://doi.org/10.6092/issn.2704-8128/12937>
- Leonetti, E., & Viale, M. (2020). La variabilità linguistica delle risposte a domande aperte di comprensione del testo nelle prove INVALSI computer based [Linguistic variability in open responses to reading-comprehension questions in computer-based INVALSI tests]. *Italiano a scuola*, 2(1), 103–130. <https://doi.org/10.6092/issn.2704-8128/10848>
- Lo Duca, M. G., & Toth, Z. (2021). La competenza ortografica nelle prove INVALSI [Orthographic competence in INVALSI tests]. *Italiano a scuola*, 3(1), 1–38. <https://doi.org/10.6092/issn.2704-8128/12931>
- Marsili, M., Campodifiori, E., Bagnarol, C., & Donno, S. (2020). *La procedura automatizzata per la codifica delle domande a risposta aperta delle prove INVALSI in modalità CBT [The automated procedure for coding open-ended responses in computer-based INVALSI tests]* (Working Paper INVALSI No. 49/2020). INVALSI. https://www.invalsi.it/invalsi/istituto.php?page=working_papers
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Noble, T., Wells, C. S., & Rosebery, A. S. (2023). English learners and constructed-response science test items: Challenges and opportunities. *Educational Assessment*, 28(4), 246–272. <https://doi.org/10.1080/10627197.2023.2226387>
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15(2), 217–262. <https://doi.org/10.1177/026553229801500204>
- Qi, Y., & Fang, Z. (2023). Teaching writing to English language learners: Two science teachers' perspectives and practices. *Journal of World Languages*, 9(3), 400–426. <https://doi.org/10.1515/jwl-2023-0016>
- Rosc-Leskovec, D. (2013). Kaj o jezikovni zmožnosti učencev kažejo dosežki nacionalnega preverjanja znanja iz slovenščine v osnovni šoli [What national Slovene assessment results reveal about students' language ability]. *Slovenščina v šoli*, 16(3–4), 16–26, 141. <https://www.dlib.si/details/URN:NBN:SI:doc-SYIMH541>
- Slepkov, A. D., & Shiell, R. C. (2014). Comparison of integrated testlet and constructed-response question formats. *Physical Review Special Topics – Physics Education Research*, 10(2), 020120. <https://doi.org/10.1103/PhysRevSTPER.10.020120>
- Smolej, M. (2021). Stanje razumljivosti in razumevanja na Slovenskem: Primer nacionalnega preverjanja znanja [The state of comprehensibility and comprehension in Slovenia: The case of national assessment]. *Linguistica*, 61(2), 31–41. <https://doi.org/10.4312/linguistica.61.2.31-41>

Received: July 22, 2026

Revised: August 06, 2026

Accepted: August 22, 2026

Cite as: Melinc Mlekuž, M. (2026). Productive minority-language use in national assessment: Implications for bilingual development. *Problems of Education in the 21st Century*, 84(4), 630–645. <https://doi.org/10.33225/pec/26.84.630>

Maja Melinc Mlekuž

PhD, Lecturer at the School of Humanities and Researcher at the Center for Experimental, Formal and Theoretical Linguistics (CEFT), University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia.
E-mail: maja.melinc.mlekuz@ung.si
ORCID: <https://orcid.org/0009-0004-8606-5776>