International Journal of Multidisciplinary Research and Growth Evaluation.

Assessing Geographical Reasoning in STEM Education through Analytic Rubrics and Formative Feedback

Viktoria Dimitrova
Faculty of Geology and Geography, Sofia University "St. Kliment Ohridski", 15 Tsar Osvoboditel Blvd., 1504 Sofia, Bulgaria

* Corresponding Author: **Viktoria Dimitrova**

Article Info

ISSN (Online): 2582-7138
Volume: 07
Issue: 06
Received: 22-08-2026
Accepted: 24-09-2026
Published: 26-10-2026
Page No: 41-45

Abstract
Background: Assessment in STEM geography needs to reveal how individual learners interpret spatial information, explain environmental processes and justify a decision within a shared project.
Objectives: This article develops a methodological model connecting these purposes with analytic criteria and opportunities to use feedback.
Methods: A targeted narrative review brings together Bulgarian geographical didactics and international research on integrated STEM education, formative assessment and rubrics. Selected published quantitative results are reported in a comparative table; no original classroom dataset is presented.
Results: The proposed model identifies five assessment criteria: process explanation, map interpretation, quantitative reasoning, justification with evidence, and evaluation and revision. An illustrative Grade 9 task on schoolyard rainwater management shows how group investigation can generate resources for individual assessment. Learners interpret a site map, calculate a scenario, compare interventions and revise an explanation after feedback.
Conclusion: The article argues that a defensible judgement requires a visible connection between the geographical learning goal, the individual response and the scoring rule. The model provides a basis for classroom adaptation and subsequent empirical evaluation; its validity and feasibility in Bulgarian secondary geography remain to be tested.

DOI: https://doi.org/10.54660/.IJMRGE.2026.7.6.41-45

Keywords: Geography education, STEM, analytic rubric, formative assessment, geographical reasoning

1. Introduction

A geography project may end with a model, a map or an environmental proposal. These products do not automatically show what each learner understands. A team presentation can contain contributions from students with different levels of geographical reasoning. Assessment therefore needs evidence that supports a judgement about an individual learner and an opportunity for that learner to improve it.

Integrated STEM education offers a setting for addressing this problem. Kelley and Knowles [1] connect situated learning with scientific inquiry, technological literacy, mathematical thinking and engineering design. In geography, a spatial or environmental question should guide the investigation. Measurements, models and spreadsheets should contribute to explaining where a process occurs, how it operates and why an intervention is appropriate.

The geographical basis of assessment can be clarified through Vasileva's model of geography education [2]. The relationship between aims, content and teaching methodology directs attention to what the lesson is intended to develop. Her distinction between cognitive, methodological, emotional and social dimensions of geographical culture also cautions against interpreting every aspect of project participation as the same outcome. The present article concentrates on individual geographical reasoning,

while recognising that collaboration and attitudes have their own educational significance.

Vasileva's discussion of competence orientation [3] further supports an alignment between intended competence, learning activity and assessment. This article translates that concern into a practical question: what must a student say, write, calculate or represent for the teacher to infer that the intended competence has been demonstrated? The aim is to develop a usable assessment model and illustrate its application in secondary geography. The contribution is methodological; it does not describe a completed classroom experiment or claim that the proposed rubric has already been validated.

2. Theoretical Background

2.1. Geographical Competence and Observable Evidence

Geographical reasoning combines the interpretation of spatial evidence with explanations of processes and relationships. A learner may identify a low point on a map without explaining why water is likely to accumulate there. Another may perform a calculation correctly while overlooking the land cover that changes its meaning. An assessment task should make these differences observable. A single overall impression of the finished product offers limited help in identifying them.

Vasileva's analysis of the map in geography education [4] distinguishes operations that progress from decoding and description towards interpretation and evaluation. This progression informs the present proposal: locating information is a starting point, while a stronger response uses mapped relationships to support a decision. Her work on geographical literacy through systems of concepts [5] also supports connecting terms within an explanation. In a water-management task, infiltration, runoff, slope and land cover should function as related concepts, rather than as a list of definitions.

These publications provide a geographical rationale for the assessment targets. They do not establish that a particular score measures geographical reasoning adequately. That interpretation requires a task which elicits the intended operations and descriptors which distinguish the quality of the resulting evidence. For example, an arrow showing the direction of flow becomes more informative when accompanied by an explanation referring to the elevation and surface information supplied in the task.

2.2. Alignment and Formative Assessment

The assessment triangle proposed by the National Research Council [6] connects a model of learning, observations that can reveal competence and the interpretation of those observations. Applied to STEM geography, the learning model identifies the geographical relationships to be understood. The task creates opportunities to demonstrate that understanding. The scoring criteria specify how the evidence will be interpreted. Each element constrains the others: a criterion about evaluating alternatives cannot be supported by a task which asks only for factual recall.

Black and Wiliam [7] place the formative role of assessment in the use of evidence to guide teaching and learning. A comment on a completed task becomes useful when the learner has a suitable opportunity to act on it. In the model proposed here, the teacher identifies a criterion, points to a gap in the response and invites a specific revision. The revised explanation then supplies evidence about the learner's response to feedback. An independent task is still needed to examine performance beyond that supported revision.

The reviews by Panadero and colleagues [8, 9] inform the use of rubrics. Jönsson and Svingby [10] distinguish scoring reliability from the validity of the interpretation attached to a score. This distinction matters when a rubric is introduced into a new subject context. Two teachers might agree on a score even if the task gives little opportunity to demonstrate the intended competence. The proposed design therefore addresses both the quality of the elicited evidence and the clarity of the rules used to judge it.

3. Materials and Methods

The article uses a targeted narrative review combined with instructional design. Sources were selected for their direct relevance to geographical competence, integrated STEM learning, formative assessment and rubric interpretation. Bulgarian geographical-didactic publications were considered alongside international research reviews and methodological guidance. This purposive selection is intended to develop a coherent classroom model; it is not an exhaustive systematic review of assessment in geography.

The synthesis proceeded from learning goals to observable responses, scoring criteria and feedback opportunities. The resulting model was examined through a constructed schoolyard rainwater-management task suitable for adaptation in Grade 9. Its maps, numerical values and possible student responses are illustrative teaching materials. No learners were recruited, no student scripts were collected and no original achievement statistics were calculated for this article.

Table 1 reproduces three estimates from the full-text Results of Panadero *et al.* [9]. No new meta-analysis was performed. The instructional proposal was developed separately from these published findings. The source evidence provides a reason to investigate the approach, while the proposed criteria and lesson sequence remain design choices requiring evaluation in their intended context.

4. Results

4.1. Published Quantitative Findings

Table 1 shows positive academic-performance estimates. The self-regulation interval includes zero. Publication-bias sensitivity analyses reduced the academic SMD estimate to 0.45 under moderate bias and 0.33 under severe bias [9]. These findings provide a foundation for the planned implementation and evaluation of the proposed approach in Bulgarian geography classrooms during the 2026–2027 school year.

Table 1: Published results for rubric use [9]

Outcome and metric	Studies (k)	Effects (m)	Estimate	95% CI
Academic performance, SMD	21	54	g = 0.57	[0.312, 0.831]
Academic performance, SMC	12	30	0.38	[0.02, 0.75]
Self-regulated learning, SMD	5	17	g = 0.23	[−0.15, 0.60]

SMD = standardised mean difference; SMC = standardised mean change; g = Hedges' g. Intervals are unadjusted for publication bias. Academic rows overlap and use different metrics. These are published findings, not author-collected student data.

4.2. A Proposed Assessment Model for Geographical Reasoning

The methodological result is a model with five criteria applied to individual responses. Each criterion is scored from 0 to 4, giving a provisional total of 20 points. Equal weighting keeps the initial design transparent, but it is not an empirically established property of geographical competence. Teachers should first decide whether all five criteria are represented adequately in the task and whether the descriptions are understandable for their learners.

The first criterion concerns geographical process explanation. A response should connect relevant conditions with a process and its consequences. In the illustrative task, the student explains how slope and surface properties influence the movement or retention of water. A response that merely repeats the word 'runoff' provides less evidence than one which traces a plausible pathway and identifies the conditions under which it may change.

The second criterion concerns map and spatial interpretation. Students use the legend, elevation information and the relative positions of features to identify a route or select a location. The response must state which mapped relationship supports the decision. The third criterion concerns quantitative and model reasoning. It includes the selection of the operation, correct use of units and interpretation of the numerical result within the assumptions of the supplied scenario.

The fourth criterion concerns justification with evidence. The learner compares possible interventions and explains why the available evidence supports a preference. The fifth concerns evaluation and revision. The student identifies an assumption or limitation and considers whether a new constraint changes the conclusion. A design can therefore be judged as well reasoned even when its recommendation differs from another student's, provided that the task permits more than one defensible solution.

The score levels follow a common progression which must be adapted to each criterion. Level 0 indicates that an available response contains no relevant evidence. Level 1 indicates an isolated relevant statement with a major error or missing relationship. Level 2 indicates a partly correct explanation with an important gap. Level 3 indicates a coherent and substantially correct response supported by the supplied evidence. Level 4 adds a justified consideration of a limitation or an alternative relevant to the criterion. Missing assessment occasions are recorded separately from a zero score.

These general levels require task-specific anchors before use. For process explanation, a level 2 response might link an impermeable surface to increased runoff but leave the route unexplained. A level 3 response would connect that surface with the slope and mapped low point. A level 4 response would also explain why a change in ground saturation or rainfall conditions could alter the outcome. Such anchors help the teacher discuss the evidence behind a judgement and give the learner a concrete direction for improvement.

5. Illustrative Application in STEM Geography

5.1. A Rainwater Management Problem

The illustrative question is how a schoolyard could retain more rainwater while preserving space for movement and recreation. The teacher supplies a schematic site map showing elevations, a building, paved and vegetated surfaces, and possible intervention locations. Teams compare a rain garden with a permeable-surface option under a stated space constraint. All site characteristics are constructed for instruction and should be replaced by appropriate local information when the task is adapted.

Science contributes explanations of infiltration and surface flow. Mathematics supports measurements, areas and volume calculations. Technology supports mapping or the organisation of observations in a spreadsheet. Engineering design enters through the comparison of interventions under constraints. Geography connects these operations through the spatial distribution of surfaces and the relationships between physical processes and human use of the site. Access to specialist software is optional; a printed map and a calculator can support the same assessment targets.

For a simple numerical scenario, 20 mm of rainfall over 100 m² corresponds to 2,000 litres of water before infiltration, storage and other losses are considered. If the task supplies a runoff coefficient of 0.70, the simplified estimate becomes 1,400 litres. Both the coefficient and the scenario values are teaching assumptions. A strong response explains what the estimate represents and why it is insufficient for predicting actual flooding without additional information.

5.2. From Group Investigation to Individual Evidence

The activity can be organised over three 80-minute sessions, with timing adjusted to the school timetable. The first session begins with a short individual interpretation of a comparable site. Students then examine the project map, discuss the assessment criteria and plan a small investigation. This initial response helps the teacher identify explanations that need attention before the teams settle on a design.

In the second session, teams compare the water collected from two model surfaces. They keep the applied water volume, surface area, slope and collection interval consistent, then describe the limitations of the model. Each learner writes an individual account connecting the observations with a map-based design choice. The teacher retains this account alongside the shared record, so that participation in the group does not become the sole evidence for an individual geographical judgement.

In the third session, feedback focuses on one criterion in the individual account. For instance, 'Explain which elevation information supports the location you selected' directs attention to a missing relationship. The learner revises the explanation and identifies the change. Students then complete an independent task using a different site and a new constraint. This sequence distinguishes supported revision from the ability to reason when the teacher's prompt is no longer attached to the answer.

The shared proposal remains useful for discussing the quality of a design and how evidence was organised. Collaboration can be observed separately through participation records and brief reflections. The individual reasoning score should be based on the designated individual responses. This separation makes it possible to recognise a student who explains the

process accurately even when the team's construction is incomplete, and to identify an unsupported explanation within an otherwise convincing presentation.

6. Discussion and Limitations

The proposed model connects the geographical purpose of an integrated task with evidence that a teacher can inspect. Vasileva's relationship between aims, content and methodology [2] provides a basis for this alignment, while her treatment of cartographic operations [4] helps distinguish locating information from using it in an explanation. The model translates those ideas into classroom assessment choices. Its contribution is this explicit connection, rather than a claim that an integrated project necessarily produces higher achievement.

A practical advantage is the possibility of directing feedback to a visible weakness. The teacher can distinguish an inaccurate process explanation from a unit error or an unsupported location choice. However, recording five scores for every classroom interaction would create an excessive workload. A feasible adaptation is to retain the complete rubric for selected individual tasks and focus on one or two criteria during formative exchanges. This choice preserves the purpose of feedback while reducing unnecessary recording.

Several limitations concern the assessment instrument itself. Written language demands may affect how well students can express their geographical understanding. A teacher should use clear prompts, provide the necessary vocabulary and check whether an oral or annotated-map response is appropriate for the intended target. Such adaptations must retain the same reasoning demand. The provisional 20-point total also needs examination: combining the criteria assumes that the summary is useful, which cannot be established from the rubric's appearance alone.

A subsequent classroom study should first examine task interpretation, descriptor clarity and scoring agreement. Two raters could independently score anonymised responses after calibration on separate examples. If an intraclass correlation is used, the model, agreement definition, single-score or average-score basis and confidence interval should be reported explicitly, following Koo and Li [11]. Disagreements at individual criteria should also be examined; a single summary coefficient cannot explain which descriptors require revision.

Evaluation should report actual participant numbers, missing observations and the conditions under which the task was implemented. For an exploratory pre-post study, paired changes and their uncertainty would be informative, alongside the time needed for assessment. Any standardised change should be identified by its calculation, since effect-size definitions answer different questions [12]. Improvement in one class would remain open to explanations such as ordinary teaching, repeated testing or task differences. A stronger causal evaluation would require an appropriate comparison design and analysis of the classroom structure.

The narrative review also has limits. Its purposive source selection can omit relevant studies, and the international assessment literature does not establish local validity for Bulgarian geography. The lesson and score descriptors are proposed applications rather than tested outcomes. These boundaries define the next research task: to determine whether the intended geographical operations are elicited consistently and whether the resulting evidence can be interpreted within the time and resources of ordinary lessons.

7. Conclusion

Assessment in STEM geography benefits from a clear connection between the learning goal, the individual response and the scoring criterion. The model developed here organises that connection around process explanation, spatial interpretation, quantitative reasoning, justification with evidence, and evaluation and revision. The rainwater-management example demonstrates how a shared investigation can support individual assessment and an opportunity to use feedback. Published research informs the rationale, while the proposed model requires its own evaluation. Classroom implementation should therefore begin with understandable criteria, appropriate response opportunities and close attention to the evidence supporting each judgement.

Use of Artificial Intelligence

ChatGPT (OpenAI) assisted with draft preparation, language editing and formatting.

References

1. Kelley TR, Knowles JG. A conceptual framework for integrated STEM education. Int J STEM Educ. 2016;3:11. doi:10.1186/s40594-016-0046-z.
2. Vasileva M. The House of Geography Education - an Example Model. Journal of the Bulgarian Geographical Society. 2019;40:59-62. Bulgarian. doi:10.3897/jbgs.2019.40.10.
3. Vasileva M. In regards of the competence orientation of the modern geography training in Bulgaria. In: 8th SWS International Scientific Conference on Social Sciences - ISCSS 2021. SGEM World Science; 2021. p. 533-540. doi:10.35603/sws.iscss.g2021/s08.53.
4. Vasileva M. The map - evergreen in geography education. Journal of the Bulgarian Geographical Society. 2019;41:94-100. doi:10.3897/jbgs.2019.41.17.
5. Vasileva M. Formation of geographical literacy through systems of concepts. In: International Scientific Conference GEOBALCANICA 2021. Geobalcanica Society; 2021. p. 593-601. doi:10.18509/GBP210593v.
6. National Research Council. Knowing what students know: the science and design of educational assessment. Washington (DC): National Academies Press; 2001. doi:10.17226/10019.
7. Black P, Wiliam D. Developing the theory of formative assessment. Educ Asse Eval Acc. 2009;21(1):5-31. doi:10.1007/s11092-008-9068-5.
8. Panadero E, Jönsson A. The use of scoring rubrics for formative assessment purposes revisited: a review. Educ Res Rev. 2013;9:129-144. doi:10.1016/j.edurev.2013.01.002.
9. Panadero E, Jonsson A, Pinedo L, Fernández-Castilla B. Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: a meta-analytic review. Educ Psychol Rev. 2023;35:113. doi:10.1007/s10648-023-09823-4.

10. Jönsson A, Svingby G. The use of scoring rubrics: reliability, validity and educational consequences. Educ Res Rev. 2007;2(2):130-144. doi:10.1016/j.edurev.2007.05.002.
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012.
12. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;4:863. doi:10.3389/fpsyg.2013.00863.

How to Cite This Article
Dimitrova V. Assessing geographical reasoning in STEM education through analytic rubrics and formative feedback. Int J Multidiscip Res Growth Eval. 2026;7(6):41-45. doi:10.54660/.IJMRGE.2026.7.6.41-45.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms