Creative Education, 2026, 17(11), 2124-2145
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

A Five-Stage Framework for Interdisciplinary Scientific Education: Integrating Archaeology, Geospatial Technologies, and Evidence-Based Research

—Lessons from Two Decades of Multidisciplinary Archaeological Investigation

Sam Osmanagich

Archaeological Park: Bosnian Pyramid of the Sun Foundation, Visoko, Bosnia-Herzegovina
Email: info@drsamosmanagich.com

How to cite this paper: Osmanagich, S. (2026). A Five-Stage Framework for Interdisciplinary Scientific Education: Integrating Archaeology, Geospatial Technologies, and Evidence-Based Research. *Creative Education, 17,* 2124-2145.
https://doi.org/10.4236/ce.2026.1711118

Received: September 11, 2026
Accepted: November 7, 2026
Published: November 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

Students in archaeology and the geosciences learn many methods but rarely learn how to combine them into one defensible argument. This paper proposes a five-stage framework that can serve as both a research protocol and a teaching model. The stages are: 1) formulation of a testable hypothesis; 2) organization of a legally authorized, institutionally continuous research program; 3) multidisciplinary investigation across four evidence streams (archaeology and laboratory analyses; geodesy, GIS and remote sensing; geophysics and environmental measurement; spatial statistics and archaeoastronomy); 4) international scientific evaluation; and 5) peer-reviewed publication. A single rule runs through every stage: no finding receives interpretive weight until it has been compared with an explicit benchmark, such as a natural control, a reference standard or a null model. Each stage pairs an educational function with defined learning outcomes, teaching activities, and assessment methods. The paper then outlines applications in university courses, field schools, and early-career research training; a one-semester curriculum with an integrated field module; and a design for future empirical evaluation. The framework draws on about twenty years of multidisciplinary investigation in the Bosnian Valley of the Pyramids. The case is used here as teaching material, including its contested interpretation. The educational model is a proposal. Its effect on learning has not yet been tested and should be evaluated independently.

Keywords

Interdisciplinary Education, Research-Based Learning, Archaeological

Methodology, Geospatial Technologies, Scientific Reasoning, Field Schools

1. Introduction

Archaeology is now a multidisciplinary science. A single project may involve excavation, radiocarbon dating, petrography, GNSS surveying, LiDAR, ground-penetrating radar, and spatial statistics (Renfrew & Bahn, 2020; Conolly & Lake, 2006). Each method answers a narrow question. None answers the larger one alone. Excavation samples a small part of a site. Geophysics detects anomalies but cannot say what they are. GIS measures spatial relationships but does not explain them. Statistics can show that a pattern is unlikely to be random yet cannot show who made it.

The hard part is integration. How should results from disciplines with different assumptions and standards of proof be combined into one argument? Kintigh et al. (2014) listed integrating heterogeneous data across disciplines as one of archaeology's grand challenges. The same problem appears in teaching. Curricula usually teach methods one at a time, in separate courses. Students learn to operate a total station or run a Monte Carlo simulation. They are seldom taught how to decide what a result means when it sits beside four other results from four other fields.

Research on higher education points in the same direction. Students gain more from authentic inquiry than from hearing about its results (Healey & Jenkins, 2009; Linn et al., 2015). Undergraduate research experiences are associated with gains in confidence, persistence and research skills (Lopatto, 2004; Russell et al., 2007). Interdisciplinary learning, however, does not happen on its own. It needs explicit structure, so that students learn to connect disciplinary insights rather than collect them (Spelt et al., 2009; Repko & Szostak, 2020).

This paper proposes such a structure. It adapts a five-stage research framework, first developed to organize a long-term investigation in Bosnia and Herzegovina, into a model for teaching interdisciplinary scientific reasoning. The framework keeps its research logic: hypothesis, organization, multidisciplinary investigation, evaluation, publication. To each stage, it adds an educational function, learning outcomes, activities, and assessment.

The framework's central rule is also its main teaching point. Every finding must be compared with an explicit benchmark before it is interpreted. The benchmark may be a natural control, a laboratory reference standard, a comparison dataset or a null model specified in advance. These benchmarks do different jobs. A laboratory reference standard establishes measurement quality: it shows that an instrument or test gives accurate, repeatable values. It does not show where a sample came from. A suitable geological comparator, such as rock from local natural outcrops tested with the same protocol, helps distinguish natural materials from manufactured ones. Students learn to ask which of the two jobs a given compari-

son does. This rule separates observation from interpretation. It also gives students a concrete habit to practice at every step.

The framework draws on about twenty years of research in the Bosnian Valley of the Pyramids near Visoko. That program is used here as a case study and a source of teaching material. The paper does not ask readers to accept the program's interpretation of the site, which remains contested (Section 7). The paper offers the educational model as a proposal. No student data are reported, and no claim is made that the model already improves learning. Section 6 sets out how that claim could be tested.

2. Conceptual Foundations

2.1. Research-Based and Inquiry-Based Learning

Healey and Jenkins (2009) distinguish teaching that presents research results from teaching in which students do research themselves. The second form is associated with deeper understanding and stronger research skills. Inquiry-based learning follows a similar logic. Pedaste et al. (2015) reviewed inquiry models and found a common cycle: orientation, conceptualization, investigation, conclusion, and discussion. The five-stage framework maps closely onto that cycle. In the Pedaste et al. model, the discussion phase comprises communication and reflection, and both can take place throughout the inquiry process, not only at its end. The framework keeps this feature. It extends the model in two specific directions: the institutional organization of research (Stage 2) and communication beyond the classroom through external evaluation and formal publication (Stages 4 and 5).

Guidance matters. Unguided discovery is inefficient for novices (Kirschner et al., 2006). Problem-based learning works best when tasks are structured, and expert support is available (Hmelo-Silver, 2004; Prince & Felder, 2006). The framework therefore works as scaffolding. It tells students what must be done at each stage and what standard the work must meet. It does not tell them what answer to reach.

2.2. Situated Learning and Field Practice

Knowledge in a discipline is learned through participation in its practice (Brown et al., 1989; Lave & Wenger, 1991). In archaeology, the field school is the main site of this participation. It is also where students learn the profession's habits of recording, sampling, and documentation (Perry, 2004; Mytum, 2012). Field schools often concentrate on excavation techniques. The framework adds the steps that come before and after: formulating the question, planning the sampling, testing results against controls, and reporting them to peers.

2.3. Evidential Reasoning and the Nature of Science

Philosophers of archaeology describe archaeological inference as reasoning from several independent, partial lines of evidence (Wylie, 2002; Chapman & Wylie, 2016). Confidence grows when independent methods converge. The same princi-

ple appears in other sciences as triangulation (Munafò & Davey Smith, 2018). Teaching this principle explicitly is part of teaching the nature of science: what counts as evidence, how hypotheses are tested, and why claims stay open to revision (Lederman et al., 2002; Popper, 2002).

Chamberlin's (1890/1965) method of multiple working hypotheses is especially useful for teaching. Students who hold only one hypothesis tend to look for confirming evidence. Students who must state several competing hypotheses, each with its own predictions, learn to design observations that discriminate between them. The framework adopts this method at Stage 1.

2.4. Constructive Alignment

Biggs (1996) argued that learning outcomes, teaching activities, and assessment tasks should align with one another. The competency matrix in Section 5 follows this principle. For each stage, it states what students should be able to do, how they will practice it, and how it will be assessed. Outcomes are written with verbs from the revised Bloom taxonomy (Anderson & Krathwohl, 2001), so that the expected cognitive level is visible.

2.5. What the Framework Adds

Inquiry cycles, research-based learning and constructive alignment are well established. The framework draws on all three. It extends them in three specific ways.

First, it carries communication into institutional organization and publication. The inquiry cycle synthesized by Pedaste et al. (2015) already includes communication and reflection, and places them throughout the inquiry process. The framework does not replace that feature; it extends it. In the inquiry cycle, communication usually stays within the class. In the framework it leads to the institutional and social stages that make research trustworthy: legal authorization and data stewardship (Stage 2), exposure to external critics (Stage 4) and formal peer review and publication (Stage 5). Students practice these stages rather than hear about them.

Second, it makes the benchmark mandatory. Inquiry models ask students to collect data and draw conclusions. They rarely specify what a result must be compared with. The framework requires a justified comparison for every finding: a natural control, a reference standard or a null model fixed in advance. Students must also argue why the chosen comparator is the right one.

Third, it teaches students to count evidence correctly. In multidisciplinary work, several analyses often rest on the same samples or the same measurements. They look like convergence but are not independent. The framework requires students to record the source of each finding and to treat several analyses of one dataset as a single line of evidence. This distinction is central to triangulation (Munafò & Davey Smith, 2018), yet it is seldom taught explicitly.

Constructive alignment supplies the structure of the teaching design (Section 5). It does not supply the content. These three features are the content.

3. The Five-Stage Framework and Its Educational Functions

The framework has five sequential stages. In research, they move a project from question to publication. In teaching, each stage also develops a specific set of competencies. Table 1 summarizes both functions.

Table 1. The five stages: scientific function and educational function.

Stage	Scientific function	Educational function	Guiding questions for students
1) Hypothesis formulation	Define testable research questions and evaluation criteria	Critical thinking and scientific reasoning	What would I expect to observe if my hypothesis were false?
2) Research organization	Secure legal authorization, institutional continuity and documentation standards	Research ethics, institutional responsibility and project management	Who authorizes this work, who is accountable and how will the data survive the project?
3) Multidisciplinary investigation	Generate independent evidence in four streams, each tested against a benchmark	Integration of archaeology, GIS, geophysics, laboratory sciences and statistics	Against what benchmark is this result being compared?
4) Scientific evaluation	Expose results to independent scrutiny at conferences and workshops	Presentation, scientific debate and peer feedback	What is the strongest objection to my result, and how do I answer it?
5) Peer-reviewed publication	Make methods and results permanently available for replication and critique	Academic writing, research transparency and scientific communication	Could another team repeat what I did from what I have written?

Note: The five stages are sequential, but findings at Stages 3 - 5 routinely return the project to Stage 1 for a revised hypothesis.

3.1. Stage 1: Hypothesis Formulation

In research, Stage 1 turns an observation into a question that can be tested. A hypothesis must state what evidence would support it and what evidence would count against it. It must remain open to modification or rejection.

The Bosnian program began with the hypothesis that several pyramid-shaped hills near Visoko were not exclusively natural landforms but contained anthropogenic components. The hypothesis concerns large-scale modification, not ordinary human use of a hill. Stated precisely, it holds that 1) the main faces of the hills were deliberately shaped into planar surfaces of regular orientation; 2) parts of these faces and of the terraces are covered with artificially produced or deliberately placed blocks and slabs; and 3) this work extends over areas of thousands of square meters on each hill, not over isolated spots. Evidence of ordinary human

activity on a natural hill does not test this hypothesis. Visočica, for example, carries the remains of the medieval royal town of Visoki, and hills in the region also show paths, agricultural terraces, small quarries and burials. Such traces are expected under the natural-landform hypothesis as well. Only observations that separate a shaped and clad hill from a natural hill with ordinary human use count as tests. This hypothesis generated questions about morphology, construction materials, chronology, subsurface structure and spatial organization. Each question could be addressed by a different discipline.

The educational function of Stage 1 is critical thinking. Students practice three skills. First, they convert a broad observation into a precise, falsifiable statement. Second, following Chamberlin (1890/1965), they write competing hypotheses. For a hill of regular shape, these would include a fully natural origin (for example, tectonic uplift and differential erosion), a natural landform with ordinary human use (settlement, fortification, terracing), a natural landform deliberately shaped and clad on a large scale, and a largely constructed feature. Third, they list, for each hypothesis, the observations it predicts and the observations it forbids.

The Visoko case is well suited to this exercise because the competing hypotheses are real and are argued in the literature. Students can be asked to write the predictions of the natural hypothesis as carefully as those of the anthropogenic one. They then identify which planned observation would best separate the two. The exercise shows that a good hypothesis is defined as much by what it rules out as by what it proposes.

3.2. Stage 2: Research Organization

Long-term multidisciplinary research needs legal authorization, stable institutions, standard documentation, and clear responsibility. Short excavation campaigns rarely produce the continuous datasets that complex sites require. In Bosnia, the Zenica-Doboj Canton issued excavation permits between 2006 and 2012. A registered foundation carried out the work, opening 20 trenches between 2005 and 2025 under continuous institutional authorization.

Research organization is seldom taught as part of the scientific method. It should be. A project without permits, sampling plans, archives and data standards cannot produce evidence that others can check. The educational function of Stage 2 is research ethics, institutional responsibility and project management.

Students learn how heritage law governs excavation and sampling. They write a short research design with a sampling strategy, a documentation protocol and a data management plan. Data management should follow the FAIR principles, so that data remain findable, accessible, interoperable and reusable (Wilkinson et al., 2016). Students also discuss obligations to local communities, landowners and heritage authorities. A practical task is to prepare a mock permit application for a defined area, including a budget, a timetable and a risk register.

3.3. Stage 3: Multidisciplinary Investigation

Stage 3 is the analytical core. Several disciplines examine the same problem using

their own methods while remaining independent of one another. The evidence is organized into four streams. In teaching, each stream becomes a training module. Every module ends with the same task: compare the result with an explicit benchmark and state what the comparison does and does not show. Table 2 lists the modules.

Table 2. The four evidence streams of Stage 3 as training modules.

Evidence stream	Core methods taught	Benchmark question students must answer	Teaching case from the Visoko program
3A. Archaeology, laboratory analyses and chronology	Excavation, stratigraphic recording, sampling design, radiocarbon and other dating, petrography, compressive-strength testing	Is the comparator appropriate? Would a local natural material be a stronger control than a modern one?	Compressive strength of 94 - 155 MPa reported for sampled blocks and compared with modern concrete (Osmanagich, 2025a)
3B. Geodesy, GIS and remote sensing	RTK-GNSS, total station, LiDAR, drone photogrammetry, DEM construction and analysis	How often does a comparable regularity occur in natural landforms of the same region?	Reported deviation of 0˚00'12" of a hill face from true north (Osmanagich, 2025a)
3C. Geophysics and environmental measurement	GPR, electrical resistivity, electromagnetic and ultrasound measurement, environmental monitoring	Were identical measurements taken on natural control hills, with the same instruments and protocols?	GPR survey reporting 44 subsurface anomalies over 10,000 m^2, and field readings compared with control hills (Korotkov & Osmanagich, 2024; Osmanagich, 2025b)
3D. Spatial statistics and archaeoastronomy	Monte Carlo simulation, permutation tests, nearest-neighbor analysis, azimuth and horizon modeling	Was the null model fixed before the data were examined? How many candidate patterns were tested?	Constrained Monte Carlo tests of spatial and astronomical hypotheses (Osmanagich, 2025c, 2026b, 2026a)

Note: The values in the last column are results from the research program and are used as teaching material. They are open to critique in class and are not presented here as settled findings.

3.3.1. Stream 3A: Archaeology, Laboratory Analyses and Chronology

Excavation establishes context: stratigraphy, features, artifacts and deposits (Harris, 1989). Laboratory work then characterizes materials and dates. Chronology

should rest on more than one dating method where possible (Aitken, 1990). Samples should follow an explicit sampling design, so that their representativeness can be judged (Banning, 2002; Orton, 2000). Geoarchaeological analysis of sediments links the two (Goldberg & Macphail, 2006).

The compressive-strength results from Visoko provide a useful exercise in benchmark choice. The reported values were compared with modern concrete. Students can ask whether that comparison answers the research question. If the question is whether the material is natural or artificial, then cemented conglomerates and sandstones sampled from local natural outcrops may be a more informative control than modern concrete. The exercise teaches that choosing a benchmark is itself a methodological decision. It must be justified, not assumed.

The same point applies to laboratory testing in general. The program reports that samples from the site have been tested at construction and materials institutes in Bosnia and Herzegovina, the Czech Republic, Slovakia, Italy and France (Osmanagich, 2025a). Such tests are not limited to measurement quality. Mineralogy, petrography, binder chemistry and microstructure can carry information on whether a material was manufactured. Their power to discriminate, however, depends on the comparison. The same tests must be applied, under the same protocol, to natural rock of the same lithology from the surrounding area. Accredited laboratory procedures make a value reliable; the geological comparator is what turns a reliable value into a test of origin.

3.3.2. Stream 3B: Geodesy, GIS and Remote Sensing

Precise spatial documentation is the basis for any landscape analysis (Conolly & Lake, 2006; Wheatley & Gillings, 2002; Kvamme, 1999). Airborne laser scanning can reveal features even under forest cover (Doneus et al., 2008). Within the framework, geospatial results are first treated as descriptions, not as inferences. A regular shape or a precise orientation is a measured variable. It is not yet evidence of design.

The reported orientation of a hill face close to true north is a good classroom case. Students can be asked how to test such a value. One approach is to measure the orientation of many natural ridges and faces in the same region, using the same DEM and the same definition of a "face". The observed value is then placed within that distribution. Students learn that a precise measurement is not the same as an improbable one.

3.3.3. Stream 3C: Geophysics and Environmental Measurement

Non-invasive methods such as GPR, resistivity and magnetometry examine the subsurface without disturbing it (Clark, 1996; Gaffney & Gater, 2003). Their results are anomalies: departures from an expected background. An anomaly is an observation, not an interpretation. Its meaning has to be established by excavation, coring or comparison with known natural structures.

Students in this module design a control. If measurements are taken on a hill of interest, the same instruments, settings and protocols must be applied to natural

hills of similar geology and size. Students also learn to report the number of anomalies relative to the area surveyed and to the background rate on control sites. Without a control, a count of anomalies has no meaning.

3.3.4. Stream 3D: Spatial Statistics and Archaeoastronomy

Apparent alignments and geometric patterns are common in any landscape with many features. They must be tested against explicit null models (Baxter, 2003; Bevan & Lake, 2013; Bivand et al., 2013). Archaeoastronomy faces the same problem. With many candidate features and many possible astronomical targets, some alignments will occur by chance (Ruggles, 2015).

This module teaches simulation and the problem of multiple testing. Students build a Monte Carlo test for a simple alignment hypothesis. They then repeat it after examining many candidate patterns and see how the false-positive rate changes. The lesson connects to current debates on reproducibility and preregistration (Munafò et al., 2017; Nosek et al., 2018). One Visoko study addresses this problem directly: hypotheses formed after inspecting the data (Osmanagich, 2026b). Students can read it critically and ask whether its constraints are sufficient. A further principle is stated explicitly: statistical significance shows that a pattern differs from a null expectation. It does not establish cultural meaning.

3.3.5. Integration across Streams

The final task of Stage 3 is integration. Students build an evidence table. Each row is one finding; the columns record the stream, the benchmark, the result of the comparison, the alternative explanations still open and the degree of independence from the other findings. Two results that depend on the same measurement do not count as two lines of evidence. The table makes this visible. It also shows where streams converge, where they conflict and where evidence is simply missing.

Table 3 gives a worked example. It uses one reported finding from each stream of the Visoko program and records it in the format students are asked to use. The entries show how the framework treats a result before any interpretation is attached to it.

Table 3. Worked example of a completed evidence table, using reported findings from the Visoko program.

Finding (observation)	Stream	Benchmark or control	Alternative explanation still open	Independence	Provisional conclusion
Compressive strength of 94 - 155 MPa in sampled blocks (Osmanagich, 2025a)	3A	Used: modern concrete (30 - 70 MPa). Stronger control: local natural conglomerate and sandstone, tested with the same protocol	Natural diagenetic cementation	Shares samples with the petrographic analyses; counts as one line of evidence with them	Does not yet discriminate between the natural and anthropogenic hypotheses; local control samples are needed

Continued

Hill face oriented 0˚00'12'' from true north (Osmanagich, 2025a)	3B	Used: published orientation deviations at other monuments. Stronger control: orientation distribution of natural ridges and faces in the same DEM	Chance orientation of a structurally controlled ridge	Independent of 3A	The measurement is precise; its probability is unknown until the regional distribution is measured
44 subsurface GPR anomalies over 10,000 m^2 (2007 survey; Osmanagich, 2025b)	3C	Needed: anomaly rate on natural control hills, same instrument and settings	Bedding planes, fractures, moisture contrasts	Independent of 3A and 3B	Observation only; meaning requires coring or excavation of selected anomalies
Spatial and astronomical patterns differing from constrained random configurations, $p < 0.05$ to $p < 0.0001$ (Osmanagich, 2025c, 2026a, 2026b)	3D	Used: Monte Carlo null models with stated constraints	Multiple testing across many candidate patterns; a null model that is too simple	Depends on the same geodetic coordinates as 3B; not independent of 3B	The patterns differ from the stated null; cultural meaning is not established by the test

Note: The entries illustrate the format and the standard the framework applies. The last column is a provisional judgment within the framework, not a final verdict on the site. In class, students complete a parallel table for the natural hypothesis.

Two lessons follow from the example. A strong-looking number may still lack the control that would make it discriminating. And four findings from four streams may amount to fewer than four independent lines of evidence.

3.4. Stage 4: Scientific Evaluation

Research is not complete when the data are analyzed. Results must be exposed to colleagues who did not produce them. Conferences, workshops and invited lectures provide that exposure. The aim is not agreement but scrutiny. Criticism identifies weak methods, missing controls and alternative explanations. Over two decades, results from the Bosnian program have been presented at international meetings and have drawn both support and strong criticism. The program has also organized nine International Conferences on Bosnian Pyramids (ICBP) between 2008 and 2026 (Section 7).

The educational function of Stage 4 is to present, debate scientifically, and provide peer feedback. Students present their evidence tables to a panel and to classmates. One group is assigned a structured role: to argue the strongest case against the presenters' interpretation. Collaborative critical discourse of this kind improves scientific reasoning (Osborne, 2010). Students also give written peer feedback using a shared rubric. Peer assessment has been shown to support learning when criteria are clear and feedback is specific (Topping, 1998; Nicol & Macfarlane-Dick, 2006).

3.5. Stage 5: Peer-Reviewed Publication

Peer review does not guarantee that a conclusion is correct. It provides an inde-

pendent check by specialists on whether methods, analyses, and interpretations meet accepted standards (Day & Gastel, 2012). Publication also makes methods and data permanent and open to replication. The author's online bibliography lists 59 scientific articles from the program as of October 2026, with the title, journal and a link for each (Osmanagich, 2026c). They cover excavation, chronology, geodesy, geophysics, laboratory analyses, environmental measurement and spatial statistics. Readers can identify each work and its venue there.

The educational function of Stage 5 is academic writing, transparency, and communication. Students write a short article in journal format. They also act as reviewers for one another, using a standard review form. Marwick (2017) has shown how reproducible workflows, with shared code and data, can be built into archaeological research. Students are asked to deposit their data and scripts with their manuscript. The final question of Stage 5 is the practical test of transparency: could another team repeat the work from what has been written and shared?

4. Educational Applications of the Five-Stage Framework in Higher Education

The framework can be used at several levels of training. The structure stays the same. The depth, autonomy, and size of the tasks change.

4.1. University Courses

At the undergraduate level, the framework can organize a single methods course in archaeology, physical geography, or environmental science. Each stage becomes a block of teaching. Students work in small interdisciplinary teams, ideally mixing students of archaeology, geodesy, geology and statistics. Each team takes one landscape question from hypothesis to a short paper. Published datasets, open LiDAR data and open satellite imagery allow much of the work to be done without fieldwork.

This design is close to course-based undergraduate research experiences (CUREs). In a CURE, an entire class participates in research with a real, unknown outcome (Auchincloss et al., 2014). The five-stage framework gives such a course a clear sequence and a shared standard of evidence.

At master's level, the framework can serve as a capstone. Students design their own project, carry out at least two of the four streams with real data, and present the results at a student conference organized as Stage 4.

4.2. Field Schools

Field schools are the setting closest to real research. They often concentrate on excavation technique. The framework widens their scope. Before fieldwork, students write hypotheses and a research design. During fieldwork, they rotate through the four streams: excavation and sampling, survey and photogrammetry, geophysics, and spatial analysis. Each rotation ends with a benchmark exercise. After fieldwork, students present results in a closing seminar and prepare a short report.

The Visoko program has run field and volunteer seasons over many years, in which participants work on excavation, documentation, and tunnel clearing. These seasons show that the logistics of a multidisciplinary field school are workable at such a site. They do not show that the five-stage model improves learning. That remains to be tested (Section 6).

4.3. Training Early-Career Researchers

For doctoral students and young researchers in research centres, the framework works as a checklist for project design and as a structure for supervision. Supervisors can review a project stage by stage. Has the hypothesis been stated with its competing alternatives? Is the program authorized and documented? Does each finding have a benchmark? Has the work been presented to critics before submission? Is the manuscript reproducible? A short written review at the end of each stage gives the candidate a record of progress and gives the supervisor an early warning of weak points.

4.4. Using a Contested Site as a Teaching Case

Contested sites are valuable in teaching because they make the rules of evidence visible. When a claim is uncontroversial, students rarely ask how it was established. When a claim is disputed, they must examine the evidence, the controls, and the reasoning on both sides. Feder (2025) uses disputed claims this way in a widely used textbook on the scientific method in archaeology.

The Visoko case can be taught as a structured debate. Students receive primary publications from the research program, published critiques and an independent geological reading of the site. The geological reading comes from a team of the Faculty of Mining, Geology and Civil Engineering at the University of Tuzla. In 2006 the team drilled six boreholes, 3 - 17 m deep, on Visočica. It described the hill as a natural formation of layered clastic sediments (conglomerate, sandstone and clay) of Miocene age, shaped by tectonic and erosional processes, and noted similar landforms elsewhere in the Sarajevo-Zenica basin (Vrabac et al., 2006). This report gives students the natural-landform explanation in the words of geologists, not only as an alternative named by the program. They build evidence tables for both the anthropogenic and the natural hypotheses. They identify which observations each side has made, which controls are missing, and what new observation would most reduce the uncertainty. The goal is not to reach a verdict in class. The goal is to practice judging evidence fairly, including evidence for a view the student does not share.

5. Learning Outcomes and Competencies

Table 4 links each stage to intended learning outcomes, teaching activities and assessment methods. The outcomes are written at the level expected at the end of a master's program. For undergraduate courses, the same outcomes can be set at a lower level of independence.

Table 4. Competency matrix: learning outcomes, teaching activities and assessment for each stage.

Stage	Intended learning outcomes (students will be able to…)	Teaching activities	Assessment methods
1) Hypothesis formulation	Formulate a falsifiable hypothesis; state at least two competing hypotheses; derive predicted and forbidden observations for each	Case analysis; hypothesis-writing workshop; multiple-working-hypotheses exercise	Written hypothesis statement scored with a rubric for testability, specificity and alternatives
2) Research organization	Explain the legal and ethical basis of fieldwork; design a sampling strategy; write a data management plan	Seminar on heritage law and ethics; mock permit application; design studio	Research design and data management plan; short oral defense
3) Multidisciplinary investigation	Apply core methods in at least two evidence streams; select and justify a benchmark for each finding; integrate findings in an evidence table	Rotating method modules; laboratory and field practicals; simulation exercises; team integration sessions	Practical method checks; benchmark justification reports; graded evidence table
4) Scientific evaluation	Present results clearly; respond to criticism with evidence; give specific and constructive peer feedback	Student conference; structured opposition role; peer review of presentations	Presentation rubric; quality of responses to the opposing team; peer-feedback quality score
5) Peer-reviewed publication	Write a manuscript in journal format; document methods for replication; review a peer's manuscript	Writing workshop; simulated peer review; deposit of data and code	Manuscript scored with a journal-style rubric; reproducibility check by another team; quality of the written review

Note: Outcomes progress from the "analyze" and "evaluate" levels at Stages 1 - 3 to the "create" level at Stage 5 of the revised Bloom taxonomy (Anderson & Krathwohl, 2001).

Three competencies cut across all five stages. The first is benchmark reasoning: the habit of asking "compared with what?" before interpreting any result. The second is interdisciplinary integration: the ability to combine results from different fields while tracking their independence and their limits. The third is intellectual honesty: reporting null results, stating limitations, and representing opposing views fairly. These three are the main targets of the model. They are also the hardest to assess, which is why each appears in more than one assessment task.

6. Implementation and Evaluation

6.1. Proposed Curriculum

Table 5 outlines a one-semester course of fifteen weeks, worth 6 ECTS credits, followed by an optional two-week field module worth a separate 3 ECTS credits. In the European Credit Transfer and Accumulation System, one credit corresponds to 25 - 30 hours of total student work (European Commission, 2015). The 6-credit course therefore requires 150 - 180 hours: 45 contact hours (three hours per week for fifteen weeks) and 105 - 135 hours of independent and team work. The field module comprises ten field days of about 7 hours each, plus preparation and a closing report, for a total of 75 - 90 hours, or 3 credits. Students who take both components earn 9 credits. The course is designed for advanced undergraduate or first-year master's students in archaeology, geodesy, geology, geography

or environmental science. Teams of four to five students should include at least two disciplinary backgrounds.

Table 5. Proposed curriculum: one-semester course (15 weeks, 6 ECTS) with optional field module (2 weeks, 3 ECTS).

Week	Stage	Topic	Main student task
1 - 2	Introduction and Stage 1	Evidence and inference in archaeology; multiple working hypotheses	Choose a landscape question; write competing hypotheses with predictions
3 - 4	Stage 2	Heritage law, ethics, sampling design, FAIR data	Research design, sampling plan and data management plan
5 - 6	Stage 3A	Excavation records, dating, materials analysis	Analyze a provided laboratory dataset; justify a benchmark
7 - 8	Stage 3B	GNSS, LiDAR, photogrammetry, DEM analysis	Measure landforms from open LiDAR data; build a regional comparison set
9 - 10	Stage 3C	GPR, resistivity, controls in geophysics	Interpret a provided geophysical dataset with control profiles
11 - 12	Stage 3D and integration	Monte Carlo methods, multiple testing, archaeoastronomy	Run a simulation test; complete the team evidence table
13	Stage 4	Scientific debate and peer feedback	Student conference with structured opposition
14 - 15	Stage 5	Academic writing, peer review, reproducibility	Submit manuscript with data and code; review another team's manuscript
Field module (2 weeks, separate 3 ECTS)	Stages 1 - 4 in the field	Rotation through the four evidence streams at an active site	Field notebook, benchmark exercises, closing seminar

Note: 1 ECTS credit = 25 - 30 hours of student work (European Commission, 2015). Weeks can be shortened or merged for undergraduate courses. Where fieldwork is not possible, open datasets can replace the field module; the course then carries 6 ECTS only.

The course can be taught by a team of instructors from different disciplines. Where that is not possible, guest sessions from specialists can cover individual streams. The structure of the framework keeps the course coherent even when teaching is shared.

6.2. Assessment

Assessment follows the principle of constructive alignment (Biggs, 1996; Biggs & Tang, 2011). Each major task is scored with an analytic rubric shared with students in advance. Rubrics improve scoring reliability and help students understand expectations (Jonsson & Svingby, 2007). A suggested weighting is: hypothesis and research design 15%; stream practicals and benchmark reports 25%; evidence table 20%; conference presentation and peer feedback 15%; final manuscript and reproducibility check 25%.

Formative feedback is given at the end of every stage, before the summative

tasks. Students revise their work in response. This cycle mirrors the research process, in which results return to earlier stages for revision. It also supports self-regulated learning (Nicol & Macfarlane-Dick, 2006; Black & Wiliam, 1998).

6.3. Proposed Design for Empirical Evaluation

The claim that the framework improves learning has not been tested. It should be tested by researchers who are independent of the Visoko program. Table 6 outlines a possible design.

Table 6. Proposed design for future empirical evaluation of the educational model.

Research question	Design	Instruments and measures	Analysis
Does the course improve students' benchmark reasoning?	Quasi-experimental pre-test/post-test with a comparison group taking a conventional methods course	Scenario-based test: students judge findings with and without appropriate controls; blind scoring with a rubric	Comparison of gains between groups, adjusting for pre-test scores; inter-rater reliability
Does the course improve interdisciplinary integration?	Same design	Scored evidence tables and written arguments; scoring by raters blind to group	Effect sizes with confidence intervals
Does the course change students' views of the nature of science?	Pre-test/post-test	Validated questionnaire on views of the nature of science (Lederman et al., 2002)	Paired comparisons; analysis of open-ended answers
How do students experience the course?	Survey and interviews	Adapted research-experience survey (Lopatto, 2004); focus groups	Descriptive statistics; thematic analysis
Do effects last and transfer?	Follow-up after 6 - 12 months	Performance on a new case study in a different setting	Longitudinal comparison

Note: The design should be preregistered before data collection. Ethical approval and informed consent are required.

A randomized design is rarely possible in university teaching. A quasi-experimental design with a comparison group and pre-tests is a practical alternative, provided its limits are stated (Shadish et al., 2002). Raters should not know which group a student belongs to. The study plan, hypotheses and analyses should be preregistered (Nosek et al., 2018). Evaluation can also follow Kirkpatrick's levels, from student reactions to learning, behavior, and results (Kirkpatrick & Kirkpatrick, 2006). The behavior level, whether graduates use benchmark reasoning in their later work, is the most important and the hardest to measure.

Courses are rarely comparable in every respect, and three differences matter most when results are interpreted. The first is students' disciplinary preparation. A class with many geodesy or statistics students may score higher on integration tasks for reasons unrelated to the framework. The evaluation should record each student's degree program and prior methods courses and use them, with pre-test scores, as covariates. The second is instructor expertise. If the framework course is taught by more experienced instructors, a gain may reflect the teachers rather than the model. Where possible, the same instructors should teach both versions,

in parallel sections or alternating years; otherwise their experience should be recorded and reported. The third is teaching time. The comparison course should carry the same credits and contact hours, and both courses should log the time spent on each topic. When several courses and instructors are involved, students should be analyzed as nested within course sections, using multilevel models (Raudenbush & Bryk, 2002), so that a difference between sections is not mistaken for an effect of the model. Differences that remain after these adjustments should still be attributed to the model with caution.

Replication at several institutions would be the strongest test. If the model works only at the site where it was developed, or only with its authors as teachers, that would be an important result in itself.

7. Discussion

The framework does not introduce new methods of research or teaching. Its contribution is organizational. It places existing methods in a sequence, adds one rule that applies at every step and pairs each step with explicit learning outcomes. The benchmark rule is the core of the model. It is simple enough to teach in the first week and demanding enough to challenge experienced researchers.

The framework's distinguishing features are set out in Section 2.5. It extends existing inquiry and research-based models in three ways: 1) it carries communication and reflection beyond the classroom, into legal organization, external evaluation and publication; 2) it requires a justified benchmark for every result, and separates calibration against reference standards from discrimination against natural comparators; and 3) it trains students to separate independent lines of evidence from several analyses of the same data. Existing models address parts of the first point. They rarely make the second and third explicit.

The framework also makes visible parts of research that students often overlook. Permits, archives, data plans, conference criticism, and peer review rarely appear in methods courses. Yet they decide whether research can be trusted. Teaching them as stages of the scientific process, not as administration, gives students a more realistic picture of science.

Using the Bosnian Valley of the Pyramids as the case study requires a clear statement, with dates. In 2006, after a field visit, Anthony Harding, then president of the European Association of Archaeologists, concluded that the formations he examined were natural (Harding, 2007). In the same year, a geological team from the University of Tuzla reached the same conclusion for Visočica from borehole data (Vrabac et al., 2006). These assessments came at the very start of the program, before most of the later work. Nearly two decades later, the 11th edition of a widely used methods textbook still discusses the site as a case of disputed archaeological claims (Feder, 2025). The interpretation of the Visoko hills as anthropogenic constructions is therefore not accepted, as of 2025, in the published positions of mainstream archaeology and geology cited here.

The research program has continued over the same period. It has organized

nine ICBP conferences (2008, 2011, 2012, 2013, 2014, 2018, 2024, 2025 and 2026). According to the organizers, more than 100 scientists from a range of disciplines and countries have taken part (ICBP, 2026). The program has also published laboratory, geodetic, geophysical and statistical results, some of them written as direct responses to criticism; one recent study, for example, deals specifically with the risk of hypotheses formed after inspecting the data (Osmanagich, 2026b). **Table 7** lists selected milestones, with their limitations and the critiques that apply to them.

Table 7. Selected chronology of the Visoko research program, 2005 - 2026, with limitations and critiques.

Year	Milestone or result	Limitation or open question	Critique or independent assessment
2005	Research program begins; initial hypothesis of anthropogenic components in pyramid-shaped hills	Hypothesis formulated from morphology	-
2006 - 2012	Excavation permits from Zenica-Doboj Canton; systematic trenching begins (20 trenches by 2025)	Excavation samples a small part of each hill	Field visit by Harding in 2006; formations judged natural (Harding, 2007); borehole-based geological report: natural clastic formation (Vrabac et al., 2006)
2007	GPR survey reports 44 subsurface anomalies over 10,000 m^2	Meaning of anomalies requires coring or excavation and comparison with control-hill rates	-
2008 - 2026	Nine ICBP conferences; more than 100 participating scientists according to the organizers	Organized by the program; not independent evaluation in the sense of Stage 4	No revised assessment from the critics cited here
2024	Environmental and electromagnetic measurements compared with natural control hills (Korotkov & Osmanagich, 2024)	Independent replication by other groups not yet reported	-
2025	Multidisciplinary evaluation: compressive strength, orientation, geophysics (Osmanagich, 2025a)	Comparators include modern concrete rather than local natural rock	Site still treated as a disputed case in methods literature (Feder, 2025)
2025 - 2026	Spatial-statistical and archaeoastronomical tests against Monte Carlo null models (Osmanagich, 2025c, 2026a, 2026b)	Coordinates shared with geodetic stream; risk of multiple testing addressed only in part	-

Note: The table lists results reported by the program and the critiques cited in this paper. It is not a complete bibliography of the program or of its critics. A dash means that no published independent assessment specific to that item is cited here.

Two points follow. First, critiques and results need dates. An early critique does not settle questions raised by later measurements, and later publications do not cancel an early critique unless they answer it directly. Second, publication of results, including in peer-reviewed journals, does not by itself confirm an anthropogenic origin. Conferences organized by the program do not replace evaluation by independent bodies. This paper does not try to settle the dispute. The frame-

work was derived from the program's experience, but its value does not depend on the program's conclusions. Its rules would apply equally to a team testing the natural hypothesis. For teaching, the dispute is an asset. It gives students a real case, with a documented history, in which the quality of controls, benchmarks and statistical tests matters, and in which they must weigh arguments from both sides.

Several limitations should be stated. First, the educational model has not been evaluated. All statements about its benefits are expectations based on the wider literature on research-based and interdisciplinary learning. Second, the model places heavy demands on instructors. Teaching four evidence streams well requires expertise or collaboration across departments. Third, real fieldwork is expensive. Open datasets reduce this cost but cannot fully replace the experience of a field season. Fourth, the framework was developed at a single site by a single research program. Its usefulness elsewhere, in research and in teaching, remains to be shown. Finally, convergence among methods raises confidence but does not remove the need for continued testing as new evidence appears.

8. Conclusion

This paper has proposed a five-stage framework for teaching interdisciplinary scientific research, built on archaeology and geospatial technologies. The stages are hypothesis formulation, research organization, multidisciplinary investigation, scientific evaluation, and peer-reviewed publication. Each stage has a scientific function and an educational function. A single rule connects them: every finding must be compared with an explicit benchmark before it is interpreted.

The paper describes how the framework can be applied in university courses, field schools, and research training. It sets out learning outcomes, teaching activities, and assessment methods for each stage, a one-semester curriculum, and an empirical evaluation design.

The framework draws on two decades of multidisciplinary research in the Bosnian Valley of the Pyramids. That program is the source of the model and of much of its teaching material. It is not evidence that the model works in education. Independent teams must now test the framework's educational value at different institutions and sites. The model is offered as a starting point for that work.

Conflicts of Interest

The author founded and led the research program used as the case study in this paper. The educational model has not been evaluated by the author or by others, and no student data are reported.

References

Aitken, M. J. (1990). *Science-Based Dating in Archaeology*. Longman.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I.

et al. (2014). Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report. *CBE—Life Sciences Education, 13,* 29-40. https://doi.org/10.1187/cbe.14-01-0004

Banning, E. B. (2002). *Archaeological Survey*. Kluwer Academic/Plenum.

Baxter, M. J. (2003). *Statistics in Archaeology*. Arnold.

Bevan, A., & Lake, M. (2013). *Computational Approaches to Archaeological Spaces*. Left Coast Press.

Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education, 32,* 347-364. https://doi.org/10.1007/bf00138871

Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University* (4th ed.). Open University Press.

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R* (2nd ed.). Springer.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5,* 7-74. https://doi.org/10.1080/0969595980050102

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher, 18,* 32-42. https://doi.org/10.3102/0013189x018001032

Chamberlin, T. C. (1965). The Method of Multiple Working Hypotheses. *Science, 148,* 754-759. https://doi.org/10.1126/science.148.3671.754

Chapman, R., & Wylie, A. (2016). *Evidential Reasoning in Archaeology*. Bloomsbury Academic.

Clark, A. (1996). *Seeing beneath the Soil: Prospecting Methods in Archaeology*. Routledge.

Conolly, J., & Lake, M. (2006). *Geographical Information Systems in Archaeology*. Cambridge University Press. https://doi.org/10.1017/cbo9780511807459

Day, R. A., & Gastel, B. (2012). *How to Write and Publish a Scientific Paper* (7th ed.). Cambridge University Press.

Doneus, M., Briese, C., Fera, M., & Janner, M. (2008). Archaeological Prospection of Forested Areas Using Full-Waveform Airborne Laser Scanning. *Journal of Archaeological Science, 35,* 882-893. https://doi.org/10.1016/j.jas.2007.06.013

European Commission (2015). *ECTS Users' Guide 2015*. Publications Office of the European Union.

Feder, K. L. (2025). *Frauds, Myths, and Mysteries: Science and Pseudoscience in Archaeology* (11th ed.). Oxford University Press.

Gaffney, C., & Gater, J. (2003). *Revealing the Buried Past: Geophysics for Archaeologists*. Tempus.

Goldberg, P., & Macphail, R. I. (2006). *Practical and Theoretical Geoarchaeology*. Wiley. https://doi.org/10.1002/9781118688182

Harding, A. (2007). The Great Bosnian Pyramid Scheme. *British Archaeology, 92,* 40-44.

Harris, E. C. (1989). *Principles of Archaeological Stratigraphy* (2nd ed.). Academic Press.

Healey, M., & Jenkins, A. (2009). *Developing Undergraduate Research and Inquiry*. The Higher Education Academy.

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review, 16,* 235-266. https://doi.org/10.1023/b:edpr.0000034022.16470.f3

ICBP (2026). *International Conference on Bosnian Pyramids* (*ICBP*): *Conferences 2008–2026*. Foundation Archaeological Park: Bosnian Pyramid of the Sun.

https://www.icbp.ba

Jonsson, A., & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review, 2,* 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. A. et al. (2014). Grand Challenges for Archaeology. *American Antiquity, 79,* 5-24. https://doi.org/10.7183/0002-7316.79.1.5

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating Training Programs*: *The Four Levels* (3rd ed.). Berrett-Koehler.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41,* 75-86. https://doi.org/10.1207/s15326985ep4102_1

Korotkov, K. G., & Osmanagich, S. (2024). Pyramids: The Influence of Form on the Environment. Part II: Bosnian Pyramids. *Acta Scientific Medical Sciences, 8,* 13-16.

Kvamme, K. L. (1999). Recent Directions and Developments in Geographical Information Systems. *Journal of Archaeological Research, 7,* 153-201. https://doi.org/10.1007/s10814-005-0002-9

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. https://doi.org/10.1017/cbo9780511815355

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. *Journal of Research in Science Teaching, 39,* 497-521. https://doi.org/10.1002/tea.10034

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate Research Experiences: Impacts and Opportunities. *Science, 347,* Article 1261757. https://doi.org/10.1126/science.1261757

Lopatto, D. (2004). Survey of Undergraduate Research Experiences (SURE): First Findings. *Cell Biology Education, 3,* 270-277. https://doi.org/10.1187/cbe.04-07-0045

Marwick, B. (2017). Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory, 24,* 424-450. https://doi.org/10.1007/s10816-015-9272-9

Munafò, M. R., & Davey Smith, G. (2018). Robust Research Needs Many Lines of Evidence. *Nature, 553,* 399-401. https://doi.org/10.1038/d41586-018-01023-3

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N. et al. (2017). A Manifesto for Reproducible Science. *Nature Human Behaviour, 1,* Article No. 21. https://doi.org/10.1038/s41562-016-0021

Mytum, H. (2012). *Global Perspectives on Archaeological Field Schools*: *Constructions of Knowledge and Experience*. Springer.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education, 31,* 199-218. https://doi.org/10.1080/03075070600572090

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The Preregistration Revolution. *Proceedings of the National Academy of Sciences, 115,* 2600-2606. https://doi.org/10.1073/pnas.1708274114

Orton, C. (2000). *Sampling in Archaeology*. Cambridge University Press. https://doi.org/10.1017/cbo9781139163996

Osborne, J. (2010). Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. *Science, 328,* 463-466. https://doi.org/10.1126/science.1183944

Osmanagich, S. (2025a). Multidisciplinary Evaluation of the Pyramid-Shaped Formation near Visoko, Bosnia-Herzegovina: A Case for Anthropogenic Construction. *Journal of Biomedical Research & Environmental Sciences, 6,* 503-529.

Osmanagich, S. (2025b). Bosnian Pyramids as Energy Amplifiers: Toward a New Understanding of Ancient Pyramid Technology. *Journal of Biomedical Research & Environmental Sciences, 6,* 683-694. https://doi.org/10.37871/jbres2122

Osmanagich, S. (2025c). Mathematical Correspondence between the Bosnian Valley of the Pyramids and the Pleiades Star Cluster: A Fibonacci-Ratio and Spatial-Alignment Study. *International Journal of Recent Advances in Multidisciplinary Research, 12,* 12003-12013.

Osmanagich, S. (2026a). A Multi-Seasonal Solar Landscape System: Cultural and Archaeological Implications of Monumental Shadow Interactions in the Bosnian Valley of the Pyramids. *Archaeological Discovery, 14,* 189-207. https://doi.org/10.4236/ad.2026.142010

Osmanagich, S. (2026b). Evaluating a Posteriori Geometric Hypotheses in Spatial Data: Constrained Logarithmic Curve Patterns in a Summit Landscape. *International Journal of Recent Advances in Multidisciplinary Research, 13,* 12229-12235.

Osmanagich, S. (2026c). *Scientific Articles [Online Bibliography].* https://www.drsamosmanagich.com/scientific-articles

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T. et al. (2015). Phases of Inquiry-Based Learning: Definitions and the Inquiry Cycle. *Educational Research Review, 14,* 47-61. https://doi.org/10.1016/j.edurev.2015.02.003

Perry, J. E. (2004). Authentic Learning in Field Schools: Preparing Future Members of the Archaeological Community. *World Archaeology, 36,* 236-260. https://doi.org/10.1080/0043824042000261004

Popper, K. R. (2002). *The Logic of Scientific Discovery*. Routledge.

Prince, M. J., & Felder, R. M. (2006). Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. *Journal of Engineering Education, 95,* 123-138. https://doi.org/10.1002/j.2168-9830.2006.tb00884.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models*: *Applications and Data Analysis Methods* (2nd ed.). Sage Publications.

Renfrew, C., & Bahn, P. (2020). *Archaeology*: *Theories, Methods and Practice* (8th ed.). Thames & Hudson.

Repko, A. F., & Szostak, R. (2020). *Interdisciplinary Research*: *Process and Theory* (4th ed.). Sage Publications.

Ruggles, C. L. N. (2015). *Handbook of Archaeoastronomy and Ethnoastronomy*. Springer.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of Undergraduate Research Experiences. *Science, 316,* 548-549. https://doi.org/10.1126/science.1140384

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Spelt, E. J. H., Biemans, H. J. A., Tobi, H., Luning, P. A., & Mulder, M. (2009). Teaching and Learning in Interdisciplinary Higher Education: A Systematic Review. *Educational Psychology Review, 21,* 365-378. https://doi.org/10.1007/s10648-009-9113-z

Topping, K. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research, 68,* 249-276. https://doi.org/10.3102/00346543068003249

Vrabac, S. et al. (2006). *Izvještaj o geološkim istraživanjima Visočice kod Visokog [Report on Geological Research of Visočica Near Visoko]*. Faculty of Mining, Geology and Civil Engineering, University of Tuzla.

Wheatley, D., & Gillings, M. (2002). *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. Taylor & Francis. https://doi.org/10.4324/9780203302392

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A. et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data, 3,* Article 160018. https://doi.org/10.1038/sdata.2016.18

Wylie, A. (2002). *Thinking from Things: Essays in the Philosophy of Archaeology*. University of California Press. https://doi.org/10.1525/9780520935402