



# Scenario-Based Teaching of Water-Stress Communication in Business Education: A Quasi-Experimental Study in Bulgaria

Teodor Daskalov<sup>1</sup>, Neviana Krasteva<sup>2</sup>, Aleksey Potebnya<sup>3</sup>

<sup>1,2,3</sup>, Faculty of Economics and Business Administration,  
Sofia University "St. Kliment Ohridski", Bulgaria

## Abstract

*Business education still lacks transferable instructional models for teaching low-salience sustainability risks that students do not readily recognize as immediate managerial and communication challenges. This paper presents and pilot-evaluates a competency-based instructional framework that uses water stress in Bulgaria as a context for teaching risk interpretation, stakeholder reasoning, and public communication before crisis conditions emerge. The Bulgarian case is instructionally relevant because official statistics show that 5.8% of the population experienced drought-related water-supply restrictions in 2024. A quasi-experimental design is implemented with two intact groups of fourth-year Business Management students. The comparison group follows an active debate-based case-learning format, whereas the intervention group completes a scenario-based sequence combining simulations, serious-game-inspired mechanisms, stakeholder role-play, and structured message mapping. The intervention is designed to strengthen engagement while integrating analysis, decision-making, and communication under uncertainty. The study examines two outcomes. First, it evaluates the quality of student discussion during structured deliberation on water-stress scenarios, focusing on argument quality, evidence use, stakeholder reasoning, and decision coherence. Second, it assesses applied communication competence through the quality of message maps and drafted public-facing statements, with attention to clarity, specificity, actionability, and trust-preserving framing. Results from the 24-protocol R1-scored corpus indicate high discussion quality in both conditions, mixed applied communication performance, and actionability as the clearest instructional bottleneck. The paper contributes an implementable teaching design, an assessment-oriented competency framework, and a transferable model for embedding sustainability-risk education in business curricula.*

*Keywords: business education; sustainability risk; water stress; scenario-based learning; communication competence*

## 1. Introduction

Water stress is not only a hydrological condition but a governance problem that becomes socially consequential through coordination, credibility, and public-facing decisions. At the global policy level, water security is also framed as directly connected to prosperity and peace [1]. In low-salience settings—where aggregate resource availability and reservoir infrastructure can foster a narrative of adequacy—capability building is often delayed until late-stage restrictions or incidents occur. This makes early-stage communication and management decisions critical: they can either stabilize expectations and mobilize preventive action or accelerate escalation through confusion, mistrust, and stakeholder conflict.

The Bulgarian case is instructionally relevant because official environmental statistics report that in 2024, 5.8% of the population experienced water-supply regimes due to drought, including 5.0% under seasonal regimes and 0.8% under year-round regimes [2]. The same source reports substantial regional variation: the most affected districts were Lovech (56.3%), Pleven (51.1%), Targovishte (33.8%), and Pernik (17.8%) [2]. The national statistic is used here as an educational anchor rather than as a full hydrological diagnosis: it helps students interpret a low-salience sustainability risk as a managerial communication problem before a crisis becomes publicly visible.

In this paper, communication-management risk under water stress is defined as the coupled risk that (a) decisions under uncertainty fail to coordinate stakeholders, allocate responsibilities, and trigger timely preventive actions, and (b) communication choices reduce trust, invite misinformation, or fail to translate decisions into actionable public guidance—thereby increasing the likelihood and cost of escalation. This definition intentionally treats communication as a governance mechanism rather than a post-hoc public-relations activity.

The educational challenge follows directly: business management curricula train managerial decision-making under uncertainty, stakeholder coordination, reputational-risk assessment, and public-facing



communication, yet water stress is often treated as a technical or environmental topic rather than a managerial communication problem. To make preparedness teachable in a low-salience context, instructional design must render latent risks cognitively salient and require assessable artefacts that demonstrate competence, not only attitudes.

The contribution of this paper is educational rather than hydrological. It offers and pilot-evaluates an artefact-based teaching design for making a low-salience sustainability risk actionable in business education. The model links scenario-based deliberation, stakeholder reasoning, decision mapping, and public-facing message production to transparent assessment criteria. The present findings are reported as pilot descriptive evidence from the complete 24-protocol R1-scored corpus and still require independent second-rater scoring before any confirmatory claim can be made.

## **2. Literature Review and Conceptual Framework**

### ***2.1 Low-Salience (“hidden risk”) Communication for Drought and Water Stress***

Slow-onset hazards such as drought and water stress are difficult to communicate because their onset is diffuse and often not readily perceived. Weitkamp et al. [3] argue that in “wet” climates drought risk may lack personal relevance and is filtered through heuristics and cultural expectations; they propose a framework for drought risk communication where publics may not perceive the risk. For education, the implication is practical: teaching must first establish interpretive frames-what to notice, what signals mean, and what actions are proportional-before competence can be expected.

### ***2.2 Water Services, Trust, and Crisis Communication as Governance***

Water services are uniquely trust-sensitive among utilities because they relate directly to health and daily life. Utility crisis communication guidance emphasizes clarity, transparency, empathy, and disciplined messaging to sustain public trust [4]. In the present framework, these principles are not treated as “soft” skills; they define performance criteria for applied risk communication: actionability, uncertainty disclosure, and decision-message alignment.

### ***2.3 Simulation and Serious-Game-Inspired Learning for Water Governance***

Scenario-based instruction is relevant to this design because it structures situated practice around realistic tasks and decision contexts in higher education [5]. Simulation and serious-game approaches are increasingly used to train water governance because they enable participants to practise coordination and decision-making under uncertainty, including stakeholder conflict and time pressure. In the water-governance literature, game-based approaches are proposed as tools to explore relational practices and complex system trade-offs [6], [7]. Recent serious-game work on drinking-water contamination emergency response explicitly models communication and multi-stakeholder coordination [8]. At European policy level, JRC guidance and training tools for water reuse governance further legitimize structured role-play and serious games as capacity-building mechanisms for risk-based planning [9], [10].

### ***2.4 Constructive Alignment, Rubrics, and Artefact-Based Assessment***

To meet education-conference expectations, the study follows constructive alignment: learning outcomes are mapped to learning activities and to assessment rubrics that score observable artefacts [11]. This assessment orientation is consistent with sustainability-competence education, which emphasizes integrated competencies and the challenge of assessing them through suitable tools [12], [13]. This alignment is important because the paper evaluates not only whether students discuss water-stress scenarios, but whether they produce evidence-aware decisions and public-facing messages that can be assessed consistently.

Analytic rubrics are appropriate for this purpose because they make criteria visible and allow separate scoring of complex performance dimensions. Jonsson and Svingby [14] conclude that rubric-based scoring can enhance reliability especially when rubrics are analytic, topic-specific, and supported by exemplars or rater training; however, rubrics do not guarantee valid judgment by themselves. Brookhart [15] similarly emphasizes that effective rubrics require criteria that match the intended performance and meaningful descriptions of performance levels. These cautions inform the pilot framing of the present



study: the rubrics are used diagnostically, while reliability and validity claims are deliberately limited at this draft stage.

**Table 1.** Competency-Artifact-Assessment Alignment

| Competency domain           | Student artefact                           | Assessment focus                                       |
|-----------------------------|--|--|
| Evidence-aware deliberation | Decision protocol, rebuttals, decision map | Argument quality, evidence use, uncertainty handling   |
| Stakeholder coordination    | Stakeholder analysis, decision package     | Interests, constraints, responsibilities               |
| Decision coherence          | Decision package and map                   | Feasibility, timeline, trade-offs                      |
| Applied risk communication  | Message map, CAP statement                 | Clarity, actionability, transparency, trust, alignment |

## 2.5 Research Questions

Because the analysis uses a pilot R1-scored protocol corpus, the study is guided by descriptive research questions rather than confirmatory hypotheses:

RQ1: What patterns of discussion quality are observable in the 24-protocol R1-scored corpus?

RQ2: What patterns of applied risk-communication competence are observable in the 24-protocol R1-scored corpus?

RQ3: Which rubric dimensions reveal the clearest relative instructional bottlenecks for future teaching iterations?

## 3. Methods

### 3.1 Design, Setting, and Instructional Timeline

We used a pilot quasi-experimental teaching design with two intact sections in a fourth-year Business Management course (total N = 30 students), organized into six teams of five students. Each section included three teams, corresponding to the comparison group and the intervention group, respectively. Sections were treated as intact groups rather than randomly assigned; within each section, teams were formed to maintain comparable team size and functional diversity. The sequence was delivered across four consecutive approximately 90-minute sessions. Both conditions shared the same core pedagogy-debate with opposing viewpoints, written decision packages, plenary rebuttal, and decision mapping-which served as an active control. The intervention condition added simulation-enhanced and serious-game-inspired mechanisms, role constraints, timed checkpoints, information updates, and structured risk-communication outputs. In this manuscript, the design is treated as a pilot comparison of artefact quality and diagnostic instructional patterns, not as a confirmatory evaluation of intervention effects.

**Table 2.** Instructional Session Timeline

| Session   | Instructional focus  | Common active-control activities  | Additional intervention elements and artefacts  |
|-----------|--|---|---|
| Session 1 | Water-stress risk framing and evidence interpretation                  | Opposing-viewpoint debate; evidence and assumptions log; initial decision package             | Scenario framing; stakeholder role cards; preliminary message map   |
| Session 2 | Stakeholder reasoning, responsibilities, and trade-offs                | Rebuttal; stakeholder analysis; decision criteria and trade-off mapping                       | Role constraints; timed checkpoint; structured stakeholder-to-message translation                                   |
| Session 3 | Decision-making under evolving information and operational uncertainty | Updated decision package; feasibility review; responsibility allocation; revised decision map | Information update round; time-pressure checkpoint; separately captured Session 3 artefacts for full-corpus scoring |
| Session 4 | Public-facing communication and decision-to-message alignment          | Final decision map; communication output; plenary comparison of decisions                     | CAP-structured public statement; source-of-truth and update-cadence requirement                                     |

Because intact groups were used, the design should be interpreted cautiously. In education research, quasi-experimental designs require attention to baseline equivalence and selection risk [16]. The



present draft-stage analysis does not claim confirmatory intervention advantages; it reports descriptive pilot evidence from the R1-scored protocol corpus.

### **3.2 Conditions: Active Control and Augmented Intervention**

Oposing viewpoints as an instructional device. The use of two deliberately opposing positions was a pedagogical technique to surface assumptions, clarify criteria, and force trade-off articulation. To minimize the risk of a false dichotomy, positions were framed as analytic poles rather than normative extremes: teams were encouraged to identify hybrid options during reconciliation and to document accepted trade-offs explicitly in the decision map. Instructors monitored for caricatured or moralizing arguments and redirected teams to evidence-based constraints, stakeholder interests, and feasibility criteria.

Comparison group (CG): debate-based case learning. Students split into two opposing positions, produced written decision packages, engaged in rebuttal, and created a decision map capturing criteria, trade-offs, and expected outcomes. Communication elements were documented but not explicitly scaffolded through role constraints, timed checkpoints, or CAP-structured public messaging.

Intervention group (IG): augmented scenario-based sequence. The intervention retained the same debate structure but added (i) stakeholder role-play through role cards with objectives and constraints, (ii) decision checkpoints under uncertainty through timed rounds and information updates, including the Session 3 update round, and (iii) explicit production of risk-communication artefacts: a message map and a CAP-structured public statement oriented to clarity, actionability, transparency, and trust preservation.

CAP refers to a Caring-Action-Perspective structure: acknowledge concern, specify actions, and provide factual context including uncertainty and update cadence.

### **3.3 Data Source and Artefact Capture**

Student protocols were collected at the end of each session and served as performance artefacts. The protocol template required evidence and assumptions, opposing viewpoints and rebuttals, stakeholder reasoning, a decision package with responsibilities, a decision map, and communication output. The full teaching sequence generated raw course materials, including session worksheets and supporting artefacts; however, the reportable corpus for the present draft comprises 24 consolidated, anonymized, and R1-scored team-session protocols from the scoring sheet. No values are inferred beyond the scored protocol records.

### **3.4 Outcomes and Rubric Scoring**

Unit of analysis. The unit of analysis is the team–session protocol (one protocol produced by one team for one session). All rubric scores are assigned at the protocol level.

Two primary outcomes were assessed using analytic rubrics applied to student protocols. O1 Discussion quality included four criteria: argument quality, evidence use, stakeholder reasoning, and decision coherence. O2 Applied risk-communication competence included five criteria: clarity, actionability, transparency/uncertainty disclosure, trust-preserving framing, and decision-message alignment. Each criterion was scored on a 0-4 scale with calibration anchors for scores 0, 2, and 4; scores 1 and 3 represented intermediate performance.

### **3.5 R1-Scored Protocol Corpus and Analysis Plan**

The analysis evaluates feasibility of artefact capture, rubric sensitivity, and diagnostic instructional patterns; it does not estimate intervention effects.

The scored corpus comprises 24 protocol-level observations and is used as a pilot artefact corpus rather than as a student-level dataset. Scores are analysed at the protocol level, and the same teams can contribute multiple protocols across sessions. The corpus therefore supports descriptive diagnostic analysis of artefact quality, but not independent-case inference or causal estimation.

At this draft stage, the rubric is used primarily as a diagnostic assessment tool to identify specific performance strengths and weaknesses, especially actionability, responsibility assignment, and decision-to-message alignment.

The instructional design generated 24 team-session observations (6 teams × 4 sessions). The R1-scored corpus comprises 24 protocol-level observations: 12 comparison-group protocols and 12



intervention-group protocols. Because the same teams can contribute multiple protocols across sessions, these observations are repeated team-level traces rather than independent cases; therefore, condition is treated only as a descriptive grouping variable.

The next strengthening step should report independent second-rater scoring, inter-rater reliability, baseline equivalence checks, and baseline-controlled comparisons where appropriate. Until that workflow is completed, the present results should be read as R1-scored pilot descriptive evidence.

### 3.6 Reliability and Scoring Workflow

The defensible workflow for stronger reporting requires two independent raters, calibration with rubric anchors, and reporting inter-rater reliability. At this draft stage, the 24 protocol-level observations were scored by one preliminary rater only. Inter-rater reliability is therefore not reported, and the findings are interpreted as descriptive pilot evidence. A second-rater procedure and inter-rater reliability estimates should be completed before the corpus is used for stronger claims. Rater blinding was not implemented at this preliminary stage; the condition labels remained visible in the working scoring file, so expectancy bias is treated as a limitation.

### 3.7 Ethics

Students were informed that protocol artefacts could be used for research in anonymized form; participation in research use was voluntary and did not affect course grading. Protocols were anonymized for research scoring, stored securely, and reported only in aggregate or through de-identified excerpts. Research scoring was separated from course grading, following informed-consent principles appropriate for educational research.

## 4. Results

Results are reported as descriptive patterns and should not be interpreted as causal intervention advantages. The condition labels are retained for transparency of artefact origin, not for estimating treatment effects. The current results report the complete 24-protocol R1-scored corpus used in this draft: 12 comparison-group protocols and 12 intervention-group protocols.

### 4.1 Pilot Descriptive Results

The Table 3 corpus comprises all 24 R1-scored team-session protocols (12 CG, 12 IG) and includes all four instructional sessions. Results remain descriptive because protocols are repeated team-level traces rather than independent cases.

The table below reports protocol-level descriptive statistics. O1 is the mean of four discussion-quality criteria; O2 is the mean of five applied risk-communication criteria. Condition statistics include *n*, arithmetic mean, sample standard deviation, and observed range.

**Table 3.** Pilot Descriptive Results and Diagnostic Interpretation (24 R1-scored protocols)

| Dimension                     | CG   | IG   | Diagnostic note           |
|-------------------------------|--|--|---------------------------|
| O1 Discussion quality         | <i>n</i> =12; <i>M</i> =3.17; <i>SD</i> =0.19; <i>R</i> =3.00–3.50 | <i>n</i> =12; <i>M</i> =3.25; <i>SD</i> =0.28; <i>R</i> =2.75–3.75 | High; small IG advantage. |
| O2 Applied risk communication | <i>n</i> =12; <i>M</i> =3.55; <i>SD</i> =0.26; <i>R</i> =3.20–4.00 | <i>n</i> =12; <i>M</i> =3.45; <i>SD</i> =0.19; <i>R</i> =3.20–3.80 | Mixed; no IG advantage.   |
| A2 Evidence use               | <i>n</i> =12; <i>M</i> =3.25; <i>SD</i> =0.87; <i>R</i> =2.00–4.00 | <i>n</i> =12; <i>M</i> =3.50; <i>SD</i> =0.67; <i>R</i> =2.00–4.00 | Modest IG advantage.      |
| B2 Actionability              | <i>n</i> =12; <i>M</i> =3.00; <i>SD</i> =0.60; <i>R</i> =2.00–4.00 | <i>n</i> =12; <i>M</i> =2.83; <i>SD</i> =0.39; <i>R</i> =2.00–3.00 | Main bottleneck.          |
| A3 Stakeholder reasoning      | <i>n</i> =12; <i>M</i> =3.75; <i>SD</i> =0.45; <i>R</i> =3.00–4.00 | <i>n</i> =12; <i>M</i> =3.83; <i>SD</i> =0.39; <i>R</i> =3.00–4.00 | High; possible ceiling.   |
| B3 Transparency               | <i>n</i> =12; <i>M</i> =4.00; <i>SD</i> =0.00; <i>R</i> =4.00–4.00 | <i>n</i> =12; <i>M</i> =3.92; <i>SD</i> =0.29; <i>R</i> =3.00–4.00 | Ceiling/near-ceiling.     |

Descriptive reporting note. *M* = mean; *SD* = sample standard deviation; *R* = observed range. Median values are omitted from Table 3 to reduce visual density; the analysis remains descriptive and protocol-level. The table is based on 24 R1-scored protocols in the uploaded scoring sheet; the R2 scoring pack is not yet available, so inter-rater reliability and reconciled scores are not yet reportable.



Criterion-level inspection indicates strong stakeholder reasoning in both conditions and a ceiling or near-ceiling pattern in transparency/uncertainty disclosure. This may reflect effective task scaffolding, but it may also indicate a ceiling effect that should be addressed in future rubric refinement. The clearest relative instructional bottleneck was B2 Actionability: students were generally able to communicate clearly and acknowledge uncertainty, but needed more support in specifying concrete actions, responsibilities, time frames, and source-of-truth references.

The descriptive results answer the research questions as follows. RQ1: discussion quality is high in both conditions, with a small descriptive IG advantage in O1 and evidence use. RQ2: applied risk-communication competence shows a mixed pattern and does not support an intervention advantage in the 24-protocol R1-scored corpus, although decision-message alignment is descriptively higher in the intervention group. RQ3: actionability is the clearest relative instructional bottleneck, especially specification of responsible actors, concrete actions, deadlines, and a source-of-truth update channel. Overall, the R1-scored pilot results support the feasibility of artefact capture and rubric-based scoring. They also indicate where the instructional design should be strengthened: not primarily in general clarity or transparency, but in actionability and decision-to-message translation under scenario constraints. No causal claims are made from the pilot corpus.

## 5. Discussion

This pilot analysis explored whether simulation-enhanced and serious-game-inspired elements produced observable diagnostic patterns beyond a strong baseline pedagogy. A key observation is that debate-based case learning with opposing positions, written decisions, and decision mapping can already elicit robust deliberation in a low-salience risk topic. This matters for business education because the task approximates managerial work: interpreting incomplete evidence, coordinating stakeholders, weighing reputational consequences, assigning responsibilities, and translating decisions into public-facing guidance.

The intervention was designed to scaffold applied risk communication through message maps and CAP-structured public statements. However, the 24-protocol R1-scored corpus showed a mixed O2 pattern rather than a consistent intervention advantage. This finding is instructionally useful because it suggests that structured communication artefacts do not automatically improve public-facing communication unless actionability, responsibility assignment, and update cadence are enforced more explicitly and practised repeatedly.

The clearest diagnostic result concerns actionability. Across the pilot corpus, students more readily produced clear, transparent, and trust-preserving messages than concrete guidance specifying who should do what, when, and through which follow-up channel. Future iterations should therefore strengthen the CAP statement template by requiring explicit next-24-to-72-hour actions, responsible actors, update cadence, and source-of-truth references.

The paper's strongest contribution is therefore diagnostic: the artefact-based rubric identifies actionability (B2) as a recurring relative instructional bottleneck, especially in the translation from transparent messaging to operationally specific public guidance.

The design-oriented value is that the assessment criteria point directly to refinements that can be implemented in the next teaching iteration before a full-corpus evaluation is attempted.

### 5.1 Practical Implications

Actionability scaffolds (immediate classroom use):

- Require every public statement to include a 24-72h action line, responsible actor(s), a deadline/review point, and a single source-of-truth update channel.
- Use a micro-checklist for students: Actor-Action-Deadline-Update channel; if any element is missing, revise before submission.

Instructors can implement the approach with minimal infrastructure by adopting four components: (1) opposing-viewpoint debate to surface assumptions relevant to managerial decision-making; (2) written decision packages with explicit responsibilities and reputational-risk implications; (3) decision mapping to force coherence, stakeholder coordination, and trade-off documentation; and (4) message mapping plus a short public statement to operationalize transparency, trust, and public-facing communication. Rubrics enable assessment without audio/video recording, reducing logistical burden.

The pilot results suggest one refinement for immediate classroom use: the communication artefact should not merely ask for a clear public message; it should require a concise action line specifying actor,



action, deadline, update channel, and next review point. This requirement would target the weakest observed O2 dimension without making the exercise substantially more complex.

### **5.2 Limitations and the Most Defensible Framing**

The full 24-observation protocol corpus is represented in the R1-scored scoring file, and no values are inferred from missing or unscored artefacts. Nevertheless, the evidence remains pilot and descriptive because the unit is the protocol, not the student, and because repeated traces can be contributed by the same teams across sessions.

The main limitations are intact-group assignment, potential diffusion between sections, protocol-based measurement that captures traces rather than full interaction, and preliminary single-rater, non-blinded scoring. Small and uneven session-level cells limit interpretation to descriptive pilot evidence and prevent inferential testing.

Opposing-viewpoint structure may introduce a false-dichotomy risk if participants interpret analytic positions as the only legitimate options. This risk was mitigated by framing positions as heuristic poles, encouraging hybrid solutions in the final decision package, and requiring decision maps to document trade-offs and feasibility constraints.

### **5.3 Future Research and Strengthening Steps**

Future iterations should ensure complete artefact capture across all teams and sessions, add independent second-rater scoring and inter-rater reliability, and incorporate delayed post-tests to examine durability. Where feasible, baseline equivalence checks and baseline-controlled analyses should be used. A component analysis could isolate which elements drive incremental gains, such as role-play, timed checkpoints, information updates, message mapping, or CAP-structured public statements. The complete protocol corpus should also use stored fidelity checklists and scoring audit notes to strengthen internal validity.

**Acknowledgement:** This article is financially supported by Project No. BG16RFPR002-1.009-0002 "Twinning2Green Transition Leadership" (TWIN2GREEN).

## **REFERENCES**

- [1] UNESCO World Water Assessment Programme. (2024). The United Nations world water development report 2024: Water for prosperity and peace. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000388948>
- [2] National Statistical Institute. (2026, February 17). Environmental statistics-2024. <https://www.nsi.bg/en/press-release/environmental-statistics-2024-8941>
- [3] Weitkamp, E., McEwen, L., & Ramirez, P. (2020). Communicating the hidden: Toward a framework for drought risk communication in maritime climates. *Climatic Change*, 163, 831-850. <https://doi.org/10.1007/s10584-020-02906-z>
- [4] Retzlaff, K. J. (2020). Water utility communications can build trust during the COVID-19 pandemic. *Journal AWWA*, 112(8), 24–31. <https://doi.org/10.1002/awwa.1553>
- [5] Almazova, N., Rubtsova, A., Kats, N., Eremin, Y., & Smolskaia, N. (2021). Scenario-based instruction: The case of foreign language training at multidisciplinary university. *Education Sciences*, 11(5), Article 227. <https://doi.org/10.3390/educsci11050227>
- [6] Magnuszewski, P., Królikowska, K., Koch, A., Pająk, M., Allen, C., Chraibi, V., & Hellman, M. (2018). Exploring the role of relational practices in water governance using a game-based approach. *Water*, 10(3), Article 346. <https://doi.org/10.3390/w10030346>
- [7] Medema, W., Mayer, I., Adamowski, J., Wals, A. E. J., & Chew, C. (2019). The potential of serious games to solve water problems: Editorial to the special issue on game-based approaches to sustainable water governance. *Water*, 11(12), Article 2562. <https://doi.org/10.3390/w11122562>
- [8] Huang, J., Ashcraft, C. M., & Mo, W. (2025). Gamifying water crisis management: A serious game for drinking water contamination emergency response. *PLOS ONE*, 20(4), Article e0321210. <https://doi.org/10.1371/journal.pone.0321210>
- [9] Allgaier, P., Maffettone, R., & Gawlik, B. M. (2026). The risk management plan challenge: A serious game on water reuse governance (EUR 40588; JRC144133). Publications Office of the European Union. <https://doi.org/10.2760/5203718>



- [10] Maffettone, R., & Gawlik, B. M. (2022). Technical guidance: Water reuse risk management for agricultural irrigation schemes in Europe (EUR 31316 EN; JRC129596). Publications Office of the European Union. <https://doi.org/10.2760/590804>
- [11] Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- [12] Brundiers, K., Barth, M., Cebrian, G., Cohen, M., Diaz, L., Doucette-Remington, S., Dripps, W., Habron, G., Harre, N., Jarchow, M., Losch, K., Michel, J., Mochizuki, Y., Rieckmann, M., Parnell, R., Walker, P., & Zint, M. (2021). Key competencies in sustainability in higher education-Toward an agreed-upon reference framework. *Sustainability Science*, 16(1), 13–29. <https://doi.org/10.1007/s11625-020-00838-2>
- [13] Redman, A., Wiek, A., & Barth, M. (2021). Current practice of assessing students' sustainability competencies: A review of tools. *Sustainability Science*, 16(1), 117-135. <https://doi.org/10.1007/s11625-020-00855-1>
- [14] Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- [15] Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, Article 22. <https://doi.org/10.3389/educ.2018.00022>
- [16] What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. Institute of Education Sciences, U.S. Department of Education. [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final\\_WWC-HandbookVer5.0-0-508.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf)