European Journal of Geography

The publication of the European Journal of Geography (EJG) is based on the European Association of Geographers' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual, or empirical way the quality of research, learning, teaching, and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.

Received: 22/10/2025
Revised: 10/04/2026
Revised: 07/07/2026
Accepted: 09/08/2026
Published: 09/09/2026

Editor:
Dr Alexandros Bartzokas-Tsiompras

DOI: 10.48088/ejg.j.mei.17.1.300.319
ISSN: 1792-1341
E-ISSN: 2410-7433

Research Article

Investigating Systems Modelling Processes in Collaborative Mystery Tasks: A Video-Based Study in Geography Education

Jennifer Meister [1], Natalie Bienert [2]✉, Annabelle Koch [3] & Rainer Mehren [4]

[1] St.-Franziskus-Schule, Olpe, Germany
[2] University of Hildesheim, Department for Geography, Germany
[3] RPTU University Kaiserslautern-Landau, Institute for Science Education, Germany
[4] Germany University of Münster, Department for Geography Education, Germany

✉ Correspondence: Natalie.bienert@uni-hildesheim.de

Abstract: Developing students' systems competence is a central goal of geography education, as it enables them to recognise, describe, and model complex human-environment systems. Previous research has mainly focused on the products of systemic reasoning, such as the quality of students' concept maps, whereas the underlying processes through which systems competence develops have remained largely unexplored. This paper therefore takes a process-oriented perspective to analyse how students with high and low levels of systems competence engage in collaborative systems modelling. The data are based on video recordings of eight student groups working on a mystery-based geography task. Using qualitative content analysis, five interrelated dimensions of systemic reasoning were derived: information processing, concept and connection formation, dealing with advanced systemic features, collaborative dynamics, and metacognitive strategies. The results reveal clear differences between high- and low-performing groups. High-performing groups demonstrate strategic information management, causal reasoning, and metacognitive monitoring, while low-performing groups show fragmented processing and linear reasoning. These findings highlight that systemic reasoning in geography is not only a cognitive activity, but also a communicative and metacognitive process. The study provides insights into the dynamic development of systems competence and offers implications for teaching, diagnostic assessment, and further research in geography education.

Keywords: systems thinking; systems competence; concept mapping; collaborative learning; learning process analysis; video-based research; content analysis; geography education

Highlights:

- Introduces a process-oriented, variance-maximising approach to systems modelling
- Identifies five interrelated dimensions of collaborative systems modelling
- High-performing groups show strategic information management and metacognitive regulation
- Low-performing groups display fragmented processing and linear reasoning

1. Introduction

In geography education, one of the core goals is to enable students to deal with the complexity of real-life socio-environmental systems. Supporting students in developing this competence is particularly important because they often struggle with complexity, including understanding interdependencies, integrating multiple perspectives, and accounting for the temporal dynamics of human-environment interactions (cf. Yoon & Hmelo-Silver, 2017). Systems competence can be understood as a key component of powerful geographical knowledge, enabling students to move beyond linear explanations and to develop more differentiated, relational understandings of human-environment systems.

In this context, mystery tasks have become an established approach. Originating from the Thinking Through Geography framework (Leat, 1998), mysteries provide fragmented information cards through which students collaboratively solve an open-ended guiding question targeting a geographical issue. Previous research shows that mystery tasks foster systemic reasoning by requiring students to identify, connect, and negotiate causal relationships, often externalised through the construction of concept maps (Benninghaus et al., 2019b; Karkdijk et al., 2019). Despite this potential, most empirical work remains product-oriented, focusing on the quality of concept maps as outcomes of systems thinking. The underlying processes through which systemic reasoning emerges in collaborative settings, particularly communicative, cognitive, and metacognitive strategies, remain largely neglected. Furthermore, although group work is widely used, comparative designs that differentiate groups based on pre-assessed levels of systems competence are scarce. Addressing these gaps requires process-focused research designs that capture the dynamic development of systemic reasoning during task performance.

The present dissertation-based study (cf. Meister, 2020) responds to this need by investigating the development of systems modelling competence during a geographical mystery task. While the underlying dissertation, published in German, primarily focused on learning strategies as its central analytical concern, the present study extends this work by employing a revised coding framework and broader analytical dimensions, in which learning strategies constitute only one component among several, thereby offering an internationally accessible perspective on findings not previously published beyond the German-speaking context. Employing a qualitative, variance-maximised design, it systematically compares high- and low-performing systems thinking groups of grade 9 students to identify process differences that would remain invisible otherwise, using video-based content analysis. Moving beyond traditional product-based evaluations, this study adopts a process-oriented perspective to further bridge the gap between observable outcomes and the collaborative, cognitive and metacognitive processes through which they emerge. In this respect, the present study is intended to complement a related product-oriented study, which instead centres on the systemic qualities of students' learning products rather than on the underlying process dynamics examined here.

2. Literature Review

2.1. Systems Thinking and Concept Mapping

Although system competence, systems thinking and systems modelling are often used interchangeably, in this study they denote distinct, albeit closely interconnected, aspects. Systems competence is an overarching construct that encompasses both cognitive and action-oriented dimensions, including the ability to analyse, model, and act upon complex socio-ecological systems (cf. Roczen et al., 2021; Mehren et al., 2018). As a core geographical skill, it comprises two dimensions: "to recognize, describe, and model a complex socio-ecological area of reality as a system" (dimension 1: system organisation and behaviour) and "to make predictions and take actions for system use and regulation based on that modelling" (dimension 2: system-adequate intention to act) (Roczen et al., 2021, p. 3). Within this framework, systems thinking refers to the cognitive processes involved in identifying system components, recognising relationships, and understanding dynamic interactions such as feedback loops and non-linear developments. Systems modelling is conceptualised as the externalised and observable manifestation of one cognitive dimension of systems competence, and thus of systems thinking. It becomes visible in activities such as constructing concept maps, where students select, organise, and connect information to represent and explain complex systems (cf. Upmeier zu Belzen et al., 2019; Schwarz et al., 2009). In this sense, modelling provides access to otherwise internal cognitive processes and allows for a process-oriented analysis of systems thinking in action.

Against this background, existing models of systems competence in geography education can be examined. While several models exist to assess systems competence in geography education, including the Systems Thinking Hierarchy Model (cf. Ben-Zvi Assaraf & Orion, 2005) and the above-cited Systems Competence Model (cf. Mehren et al., 2018; Roczen et al., 2021), they primarily focus on assessing outcomes and products, such as concept maps. However, these models pay limited attention to the underlying processes involved in systems thinking, such as concept map construction as a form of systems modelling (dimension 1). This pattern is also reflected in recent geography education research. A systematic review of concept map use in geography education shows that concept maps are predominantly employed as assessment tools to reveal conceptual change, rather than to analyse the processes through which students construct systemic understanding (Pivarníková, 2025). The review further highlights that concept mapping can support the development of systems and relational thinking while helping to identify misconceptions, underlining its potential as a learning tool (ibid.). This aligns with recent research in science and geography education, in which systems thinking should be conceptualised not only as an internal cognitive ability, but as a process that unfolds through modelling activities in which learners externalise and iteratively refine their understanding of complex systems (Upmeier zu Belzen et al., 2019; Schwarz et al., 2009; Nielsen & Nielsen, 2021). Recent modelling research also suggests that successful modelling depends on more than the final product alone and involves further dimensions such as modelling practice and meta-modelling knowledge (Göhner et al., 2022).

Empirical evidence suggests that students who actively construct and revise concept maps develop deeper conceptual understanding, improve their ability to integrate multiple system components, and are better able to deal with complexity (Lenski et al., 2022; Cañas et al., 2023). For example, Cox et al. (2022) show that concept maps are effective tools to visualise students' knowledge structures and to assess the complexity and interconnectedness of their understanding. However, the analysis remains focused on the structure and quality of the resulting maps, rather than on the processes through which these representations are constructed.

Concept-mapping procedures are applied in geography lessons within the framework of the mystery method, which has gained increasing attention in both classroom and research. Developed within the Thinking Through Geography-Approach (cf. Leat, 1998), these open-ended learning activities revolve around students working collaboratively to solve a guiding, problem-oriented question by using fragmented, discontinuous information that is typically presented on cards and may include text, images, diagrams, statistical data, and/or maps (cf. Leat & Nichols, 1999). Students are encouraged to identify, select, and connect relevant information to form explanations regarding the guiding question through visualisation in a concept map. Because of the open-ended guiding question that cannot be answered with single, linear reasoning and the fragmented information, mysteries require students to engage in extensive discussion and negotiate their concept map construction process (cf. Leat, 1998; Leat & Nichols, 1999). In the Thinking Through Geography approach, collaborative group work plays a central role. High-quality groups are characterised by reciprocity and speculation. Reciprocity refers to active listening as well as joint systems modelling. Speculation involves the formulation of assumptions – for example, in the form of what-if questions (cf. Leat, 1998). Recent research on systems competence emphasises that collaborative interaction and linguistic negotiation processes are central to the development of systems competence, as learners co-construct explanations and negotiate meanings through discourse, thereby achieving a deeper understanding of the system than would be possible without collaboration (Kwon & Cifuentes, 2009; Chang et al., 2017; Schuler et al., 2018). Demssie et al. (2023) reach a similar conclusion, stating that collaborative approaches strengthen learners' systems thinking through the sharing of ideas and assumptions.

Mysteries therefore offer considerable potential for systems thinking on various aspects. First, they support students in recognising interdependencies and moving beyond monocausal explanations (cf. Karkdijk, 2021). Second, they require relational and multi-perspective thinking, both of which are central elements of systems competence (cf. Mehren et al., 2018). Third, higher-order systemic reasoning processes are induced through stimulation of dialogue and collaboration in small group settings (cf. Benninghaus et al., 2019a; Lamb, 2009). Finally, by integrating multiple causes and effects, mysteries foster holistic explanations of human-environment systems, making them particularly suitable for sustainability-related geographical issues (cf. Benninghaus et al., 2019b). In this regard, systems competence is also considered a key competence for addressing sustainability challenges, as it requires dealing with complexity, uncertainty, and multiple perspectives (Demssie et al., 2023). Given these theoretical advantages, mysteries are not only a promising tool to promote systems thinking in geography classrooms, but also a valuable research tool: "Mysteries can be part of regular classroom instruction and can be used subsequently for analysis" (Benninghaus et al., 2019b, p. 722).

2.2. Empirical Results on Mystery-based Concept Mapping Tasks

Concept mapping has recently gained increasing attention as a tool for investigating students' systems thinking in geography education. In the vast majority of the available studies, concept maps are used as intervention and diagnostic tools (Khajeloo & Siegel, 2022; Mehren et al., 2018; Hempowicz, 2021; Cañas et al., 2023; Tripto et al., 2017; Ben-Zvi Assaraf et al., 2013; Li & Li, 2023; Benninghaus et al., 2019a). The following section reviews process-oriented findings on mystery-based concept mapping. Based on a multi-method research design – consisting of photographed student products, video data, stimulated-recall interviews and observational notes – Leat and Nichols (2000) identified five phases of mysteries. In the display stage, information cards are reviewed and sorted, followed by the setting stage, in which a thematic grouping of the cards takes place. High-achieving learners drew upon implicit background knowledge and made more nuanced categorisations, whereas lower-achieving learners drew premature conclusions and discarded potentially relevant cards (Leat & Nichols, 2000). In the sequencing and webbing stage, causal links are established between the information cards, whereby both linear and complex, non-linear connections can arise. The reworking stage comprises the reorganisation and restructuring of information, whereby a higher frequency of restructuring is associated with an improved quality of the learning product. In some groups, an abstract stage can also be observed, in which learners discuss the content without physical manipulation of the cards.

Building on this work, Karkdijk et al. (2019) adopted a comparative perspective in a video-based study involving 35 triads (aged 15–17) to explore the students' systems thinking. Using the phases proposed by Leat and Nichols (2000) together with the SOLO taxonomy (Biggs & Collis, 1982), the authors developed a deductive coding framework that further differentiated the individual phases. High-performing groups were characterised by the causal linking of information cards during the sequencing and webbing stage, whereas lower-performing groups predominantly made linear or chronological arrangements and more frequently classified general information cards as irrelevant. Regarding collaborative cooperation, it was evident that high-performing groups exhibited more intensive cooperation involving task-related discussion. Although the comparative design explored valuable contrastive insights, the distinction between the groups was based on the concept maps created using the criterion map. However, group composition was not determined prior to data collection.

Growing interest in process-based empirical research is exemplified in a study conducted by Hempowicz (2021). In this qualitative case study, he investigated group work (n=3 groups) on mystery-based concept mapping in geography by combining pre-assessments of students' system organisation competence with video analyses of their collaborative processes. The more intensive cooperation of high-performing groups found by previous research, however, was not replicated. Instead, Hempowicz found that the collaboration within groups of three was uneven, requiring further empirical clarification. Overall, the available evidence suggests that the quality of mystery processing is significantly determined by the cognitive strategies used as well as the intensity of collaborative interaction.

Recent research in modelling and science education also indicates that metacognitive processes and modelling knowledge are central for understanding how students engage in complex systems. Studies show that students need to actively plan, monitor, and adapt their modelling strategies to develop coherent explanations and integrate multiple system components (Ammoneit et al., 2024).

3. Aims and Research Questions

Three main methodological gaps remain in research on mystery-based concept mapping and systems thinking: First, most studies focus on product-based outcomes, such as concept maps, rather than the processes through which students construct systemic understanding (cf. Khajeloo & Siegel, 2022; Li & Li, 2023). Addressing this gap is essential for advancing students' systems competence and supporting formative assessment (cf. von Aufschnaiter et al., 2018). Second, comparative approaches in group work are rare: systems thinking is often explored individually, but collective construction remains largely unexplored. Even when addressed, studies do not systematically differentiate between pre-assessed levels of systems thinking (cf. Karkdijk et al., 2019), are based on small case studies (cf. Hempowicz, 2021). Moreover, findings on the role of collaboration remain inconsistent (cf. Leat & Nichols, 2000; Hempowicz, 2021). Third, metacognitive processes during systems modelling remain underexplored, with little attention paid to the reflective strategies that underlie systemic thinking (cf. Hempowicz, 2021; Ammoneit et al., 2024; Zohar & Dori, 2012). Addressing these gaps requires process-focused, comparative designs that systematically investigate how systems competence unfolds in collaborative concept mapping tasks. To address these gaps, the present study adopts a qualitative, variance-maximised design to investigate systems modelling processes during collaborative concept map construction in a geographical mystery task. Rather than evaluating the effects of an intervention, it contrasts groups with high and low levels of systems competence. This leads to the following research questions:

- RQ 1: How does the process of collaborative concept map construction in a geographical mystery task differ between high- and low-performing systems thinkers?
- RQ 2: Which differential categories can be identified between those groups?

To address these research questions, typical process patterns of systemic modelling are first identified during concept map construction. These patterns are then compared across high- and low-performing groups and subsequently discussed in light of existing research.

4. Materials and Methods

Drawing on a doctoral dissertation (Meister, 2020), this study explores how groups of students with different levels of systems competence engage in collaborative concept map construction within a mystery task, focusing on the cognitive, metacognitive and collaborative dynamics involved in the development of systems modelling competence. Figure 1 summarises the study design.

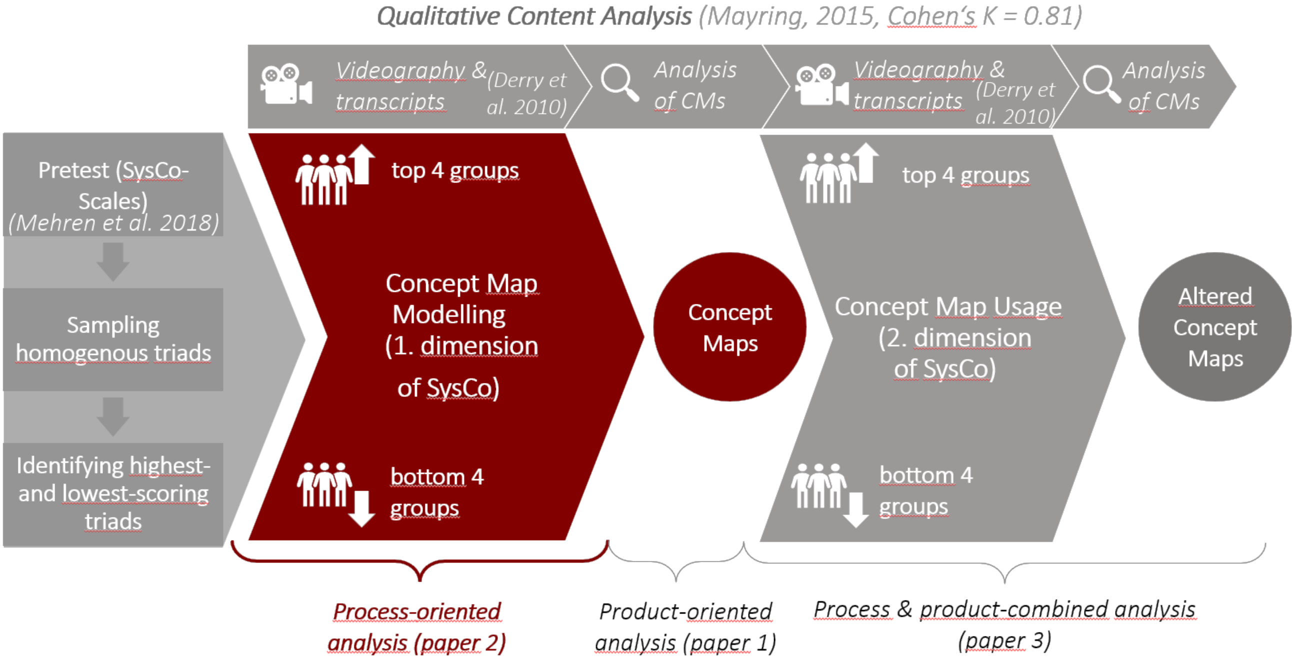

Figure 1. Research flowchart (illustration: authors).

4.1. Sample

To maximise variation, participants (n=30) were recruited from different school types (comprehensive school, secondary school) in North Rhine-Westphalia in year 9 (i.e., aged 14-15), equivalent to lower secondary education (e.g., year 9 in the UK).

To minimise the influence of differing levels of systems competence and to allow for a theory-based comparison of group behaviours, students were grouped into homogeneous triads with either low or high systemic competence based on a pre-test in a variance-maximising approach (cf. Kleemann et al., 2009; Fig. 1). This pre-test was based on a validated systems competence test (Mehren et al., 2018) and included tasks requiring students to identify system components, describe relationships between elements, and reason about dynamic interactions within systems. Responses were evaluated using predefined criteria reflecting increasing levels of systems competence (e.g., linear, relational and dynamic reasoning) on the dimensions of systemic modelling and system-adequate intention to act. Students were then assigned to either high-or low-performing groups to maximise between-group variation. Participants were not informed of their group allocation to prevent stigmatisation and to avoid influencing the data collection. Furthermore, the system competence test had no impact on their school grade, allowing them to engage without the usual pressure associated with grading. In all, four high-performing and four low-performing systems competence triads were formed (see Fig. 1).

Data collection took place during regular geography lessons in 2016, ensuring that the study was embedded in authentic classroom practice rather than an experimental setting. Because students within each triad were familiar with each other from regular classroom interaction, this reflects typical classroom conditions and supports ecological validity,

https://doi.org/10.48088/ejg.j.mei.17.1.300.319

but may have influenced collaborative dynamics and is addressed in the discussion. All groups were given a standardised, paper-based mystery task on a geographical system, centred on the guiding question “Why are there so many pirates in Somalia?”. As with any context-specific learning task, students may have differed in their prior knowledge and media exposure regarding piracy in Somalia. However, the mystery deliberately and explicitly addressed the then highly mediatised topic from multiple perspectives to move beyond simplified or stigmatised explanations: The task comprised 26 information cards involving aspects such as environmental, economic, and political issues of the Somali piracy problem. A standardised instructional setting ensured comparable conditions: students were instructed to collaboratively analyse the cards, sort and connect the cards, construct a concept map on a large sheet of paper and present their explanation to the guiding question based on their concept map.

4.2. Instruments and Data Sources

In line with the research aim of analysing the development of systems thinking, videography was chosen as the primary data collection method because it allows learning processes to be captured in detail (cf. Derry et al., 2010; Tuma et al., 2013). During the 45-minute mystery task, each group's work was video-recorded using two stationary cameras to capture verbal interactions, gestures, concept map development, and the use of task materials. This resulted in an audiovisually rich dataset comprising eight group recordings (see Fig. 1). All recordings were transcribed verbatim for analysis. In line with previous research, particular attention was paid to students' verbal interactions, as linguistic negotiation is considered an important indicator of collaborative systems modelling processes. Data triangulation was achieved using scanned concept maps, researchers' field notes, and observation logs documenting the timing and sequence of students' actions, increasing reliability and validity through triangulation of data. Together, these data sources provide a comprehensive dataset for understanding how systems competence unfolds dynamically in collaborative settings.

4.3. Content Analysis

4.3.1. Coding System

The data were analysed using a deductive-inductive qualitative content analysis framework (cf. Mayring, 2015). The coding system was developed in a multi-stage process based on the underlying doctoral dissertation. In an initial inductive phase, a broad set of 42 categories was generated directly from the video data, capturing observable modelling strategies and actions during the modelling process. These categories were subsequently reduced to 25 categories representing the greatest contrasts between high- and low-performing groups. In a further step, these categories were analytically condensed and theoretically derived. For example, statements such as “should I read all cards first?” were summarised under the category “structured reading”, which had already been addressed by Leat & Nichols (2000) in the display stage.

The initial coding followed an inductive approach, in which descriptive labels (e.g. "structured reading") were assigned to observed actions. The subsequent categorisation aligned these codes with the theoretically derived analytical dimensions (e.g. “information processing”), which were based on three theoretical and empirical layers: research on systems thinking (cf. Ben-Zvi Assaraf & Orion, 2005; Mehren et al., 2018), cognitive and metacognitive operations, and collaboration (cf. Leat & Nichols, 2000; Karkdijk et al., 2021). Drawing on the literature review in section 2, the analytical framework of this study thus assumes that systems thinking becomes visible in the interaction of cognitive, metacognitive, linguistic, and collaborative processes (see Table 1 and Section 2). As systems thinking remains partly internal, its development can only be reconstructed through observable modelling practices such as concept map construction, verbal reasoning, negotiation, and strategic regulation during task completion. Accordingly, five analytical dimensions were derived to capture the process of systems modelling competence in collaborative mystery work: 1) information processing (i.e., sequencing and categorising information cards, identifying relevance), 2) concept and connection formation (i.e., generating explanations, causal reasoning), 3) dealing with advanced systemic features (i.e., physical structuring of the concept maps, relations), 4) collaborative dynamics (i.e., questioning, justifying, revising ideas), and 5) metacognitive strategies (i.e., monitoring and reflecting).

https://doi.org/10.48088/ejg.j.mei.17.1.300.319

Table 1. Analytical Dimensions and Theoretical Foundations

Analytical Dimension	Description	Theoretical Foundation
Information processing	Selection, structuring, and organisation of information cards	Cognitive structuring of knowledge and modelling as an process of organising information (corresponding to display stage (Leat & Nichols, 2000; Karkdijk et al., 2021), system organisation (Mehren et al., 2018), ability to identify components and processes of the system (Ben-Zvi Assaraf & Orion, 2005))
Concept and connection formation	Development of concepts and explicit relationships between system components	Concept formation as a central mechanism of learning; modelling through relational structuring (corresponding to system structure and interaction (Mehren et al., 2018), setting stage (Leat & Nichols, 2000; Karkdijk et al., 2021), ability to identify relationships among components (Ben- Zvi Assaraf & Orion, 2005))
Dealing with advanced systemic features	Engagement with complexity, feedback loops, uncertainty, multiperspectivity, and temporal dynamics	Systems thinking as understanding non-linear, dynamic systems and future-oriented reasoning (corresponding to interaction and dynamics (Mehren et al., 2018), sequencing and webbing stage (Leat & Nichols, 2000; Karkdijk et al., 2021), ability to identify dynamic relationships within the system (Ben-Zvi Assaraf & Orion, 2005))
Collaborative dynamics	Interaction patterns, participation, discourse structure, and co-construction of meaning	Learning as a social, dialogic process; knowledge construction through interaction and discourse (corresponding to abstract stage (Leat & Nichols, 2000; Karkdijk et al., 2021), see also Hempowicz, 2021)
Metacognitive strategies	Planning, monitoring, evaluating, and adapting modelling processes	Metacognition as regulation of thinking and learning processes; modelling as self-regulated activity (corresponding to Ammoneit et al., 2024; Zohar & Dori, 2012)

While the overarching dimensions are theoretically informed, the coding system also reflects the empirical structure of the data, ensuring both theoretical grounding and sensitivity to the observed learning processes. Each category was defined by explicit coding rules and linked to illustrative examples from the dataset. A detailed coding scheme, including definitions and example excerpts, is provided in the appendix.

4.3.2. Coding Procedure

Applying these categories enabled a comparison of systems modelling processes between high- and low-performing groups. The analysis followed the four-step procedure proposed by Mayring (2015) and was conducted using MAXQDA 2020. First, to ensure a systematic analysis, video transcripts were segmented into meaning units representing coherent actions or reasoning sequences rather than isolated utterances. This step was followed by open coding, in which the segments were labelled with low-inference descriptors, such as "participant suggests causal link" or "participant challenges group decision". The coded segments were then categorised according to the main categories outlined above. Finally, coded segments were comparatively analysed across the two group types (high vs. low performers). The comparative analysis systematically identified similarities and differences between high- and low-performing groups, enabling the reconstruction of typical process patterns associated with different levels of systems modelling competence (cf. Mehren et al., 2018). The resulting differences of high- and low-performing groups were presented in comparative profiles and illustrated with transcript excerpts. Selected excerpts were also compiled as short video vignettes for use in pre-service teacher education and professional development.

Several measures were taken to ensure reliability and validity. First, two coders independently coded 30% of the dataset, resulting in a high inter-rater agreement (Cohen's K=0.81). Coding disagreements were discussed to reach a consensus. Weekly peer debriefings with an interdisciplinary research team and reflexive memoing strengthened the trustworthiness of the analysis (Lincoln & Guba, 1985). In addition, data triangulation was applied by combining video recordings, transcripts, students' concept maps and observer protocols, enabling cross-validation of identified process patterns and strengthening the validity of the findings. For example, the transcript data did not always indicate which cards the learners were specifically referring to. In these cases, the video recordings were consulted, as gestures (e.g. pointing or tapping) were frequently made during the conversation. Combining both data sources strengthened validation of the interpretation.

5. Results

The video-based content analysis provides qualitative insights into differences in systems modelling processes between high- and low-performing groups working on a human–environment system. Particular attention was paid to recurring patterns of interaction, reasoning and strategy use during collaborative work on the mystery task. Across all groups, a range of systems modelling strategies and actions was observed. These were systematically analysed and grouped into higher-order dimensions that form the basis of the following results.

The analysis identified five interrelated dimensions of systems modelling processes. Information processing describes how students selected, organised, and structured information. Concept and connection formation refers to the generation of concepts and causal relationships. Dealing with advanced systemic features captures the recognition of dynamics, uncertainty, and complexity. Collaborative dynamics describes patterns of interaction and argumentation, whereas metacognitive strategies include planning, monitoring, and reflection during systems modelling. These dimensions represent recurring patterns across all groups and capture cognitive, metacognitive and collaborative aspects of systems modelling and can be understood as the central analytical categories that differentiate systems modelling processes between high- and low-performing groups (see Table 2).

Table 2. Differences between high- and low-performing systems thinking groups across the five analytical dimensions

Dimension	High-performing groups	Low-performing groups
Information processing	Planned & structured	Spontaneously & selectively
Concept and connection formation	Reflected & precise	Unstructured
Dealing with advanced systemic features	Multi-perspective, hypothetical & dynamic	Linear, monocausal & static
Collaborative dynamics	Collaborative & dialogic	Unbalanced & monological
Metacognitive strategies	Proactive, reflective & adaptive	Reactive, reproductive & less reflective

Before presenting the comparative analysis in detail, a descriptive overview of each dimension is provided to illustrate the range or observed strategies across all groups. Additionally, selected frequency-based observations are included to provide descriptive orientation. However, they are not intended as statistical measures or inferential evidence but as indicators of recurring qualitative patterns across the analysed groups. Following Kuckartz and Rädiker (2022), reporting category frequencies in qualitative content analysis helps to illustrate how consistently particular patterns occurred across cases and thereby supports, rather than determines, the qualitative interpretation.

5.1. Information Processing

Clear differences between high- and low-performing groups' application of information processing strategies were observed. While high-performing groups showed strategic and coordinated approaches to information processing, low-performing groups tended to display ad hoc and reactive patterns of engaging with information. These differences became apparent within the first five minutes in their initial approach to the available information. High-performing groups (e.g., G3, G5) began the mapping task with a strategic reading of the mystery cards, which they read and reviewed completely before starting to sort them thematically. Group 8's comment exemplifies this strategic approach clearly: "Should I read these out loud [holding all mystery cards]?" (p.191). They also highlighted important content through colour-coding, i.e. using red marker for causes and blue marker for effects: "Let's use different colours so it's more well-arranged" (group 6, p. 191), thereby establishing a shared overview and coordinating group understanding before engaging in deeper reasoning. In contrast, low-performing groups' approaches differed markedly from this, showing a more fragmented and unsystematic approach to information intake. Rather than establishing an overview, they often read the cards spontaneously and had ad hoc discussions about individual cards, without prioritisation or integration into a coherent structure.

Beyond these organisational differences, further differences arose regarding the topic focus and problem definition. Early on, high-performing groups started to form their own hypotheses about causes of the systemic problem "Piracy in Somalia". In contrast, low-performing groups tended to seek simple, monocausal explanations: "There are so many pirates in Somalia because back then, there used to be many pirate fights" (group 10, p. 235). The analysed discussions suggest different approaches to the problem: while high-performing groups more frequently constructed multidimensional explanations, low-performing groups tended to rely on more linear explanatory patterns.

5.2. Concept and Connection Formation

Across all groups, concept and connection formation was characterised by the formulation of explanations, the establishment and validation of causal relations, and the integration of various information elements. Main differences between the groups were observed in the way these connections were established and justified. While low-performing groups primarily repeated the concepts on the information cards without further abstraction or introducing additional concepts, high-performing groups identified new concepts, such as "failed government" (group 8) in their discussion, indicating a high level of abstraction and integration of previous knowledge. These concepts also appeared frequently in the concept map itself, indicating semantic elaboration.

Major differences were also observed regarding the relationships between concepts, with high-performing groups often explicitly discussing the type of connection:

> *"S1: "Yeah, let's just write 'exchange' at the top, and here they also have something to do with piracy, because it's shipped by boats {/points to the trade group/} (.) like, stuff can be shipped to Asia and then the pirates attack it {/points to the piracy cluster/}."*
>
> *S2: "Oh right, so that supports trade and actually also supports piracy, because they live off what they steal {/points from trade to piracy/}."*
>
> *S3: "Yeah, or they attack, they can do both {/points to piracy cluster/}."*
>
> *S2: "Yeah, we can do both connections, that works." (group 6, p. 196)*

Moreover, they introduced and discussed potential alternative connections (group 5). In total, these groups frequently engaged in explicit reasoning processes, in which proposed connections were discussed, justified, and further refined within the group. Their discussions contained a greater number of negotiated alternative connections and references to more complex systemic relationships. Low-performing groups' approaches differed markedly from this, formulating connections in a more implicit or linear manner through simple statements, with the type of connection remaining unspecified: "K10 leads to K14" (group 7b, p. 169). As this illustrates, relationships between concepts were often stated without further elaboration or remained limited to simple cause-and-effect chains. Explanations, when given, were also less frequently negotiated within the group and often remained at a descriptive level, without considering more complex systemic interactions. All in all, the analysed discussions suggest that the groups differed not only in the number of connections established, but also in how these connections were articulated, justified, and negotiated during the modelling process.

5.3. Dealing with Advanced Systemic Features

Students used systemic features to varying degrees in order to go beyond linear cause-and-effect relationships. Main differences between the groups became apparent in the extent to which these systemic features were identified and integrated into the modelling process. High-performing groups more frequently identified complex interdependencies and incorporated dynamic aspects into their explanations, often going beyond linear causality and reflecting a more holistic understanding of the system. Illustrating this, dynamic, circular system properties were exclusively recognised and referred to by high-performing groups. On average, high-performing groups referred to systemic feedback dynamics approximately 4 times per discussion, whereas low-performing groups did not refer to them at all. Evidently, these groups rarely addressed advanced systemic features explicitly, with explanations focusing on isolated cause-and-effect relationships without considering feedback processes or temporal dynamics. Complex relationships were either simplified or overlooked, resulting in more static and fragmented representations of the system. For example, low-performing groups mainly remained on a level of "This leads to that" (group 9, p. 176), while high-performing groups' utterances are marked by higher complexity, such as "somehow different things, which you'd actually think have nothing to do with each other, are still connected - because I mean, like the fishing [points from 'population' to 'terrorism'], it doesn't really have much to do with terrorism, but in the end, it's all a circle" (group 6, p. 168), pointing towards an advanced understanding of circular systemic features.

Moreover, the groups also differed in how they dealt with uncertainty. Low-performing groups stuck closely to the given information and expressed little to no uncertainties: "Just glue it down and draw an arrow" (group 7, p. 169). On average, they discussed alternative relationships less than once per discussion. In contrast, high-performing groups discussed alternatives an average of approximately 5 times per discussion, articulated doubts and additional explanations: "Maybe it's because Somalia has a lot of natural resources or something important, and those get shipped away — so the pirates can then raid the ships." (group 6, p. 166). Tying in with this, they also frequently brought up different scales and perspectives, such as the views of the EU vs. locals, and integrated hypothetical thinking.

All told, clear differences between the groups became particularly visible in this dimension, with high-performing groups demonstrating an emerging understanding of systemic complexity, while low-performing groups largely remaining at the level of simplified linear reasoning. The findings suggest that engagement with uncertainty, feedback loops, and multiple perspectives constituted a key distinguishing feature of the analysed groups.

5.4. Collaborative Dynamics

The collaborative dynamics were characterised during the modelling process by interaction patterns, participation patterns and communication structures, including the distribution of speaking turns, responsiveness towards group members, and the argumentative structure within the group. The groups differed in the extent to which these collaborative dynamics were realised. First, the participation of speakers in the triads varied between the groups: in all high-performing groups, all three students actively participated in the discussion as speakers. In contrast, there were only one or two active speakers in low-performing groups, indicating a pronounced imbalance in active speaking participation and markedly different participation structures during collaborative knowledge construction.

Beyond these numerical differences, the groups also differed in the overall discussion structure. On the one hand, high-performing groups' discussions were marked by collaborative dynamics, manifesting in frequent turn-taking, references to previous speakers' comments (i.e., group 8: "Wait - she said something smart, we also order stuff online that gets shipped close to Somalia", p. 209) and metacognitive comments (see below). These interaction patterns consequently indicate a high degree of responsiveness and joint meaning-making, in which ideas are taken up, elaborated, and integrated into the group discourse. On the other hand, low-performing groups tended to have one member dominating the discussion (i.e., group 7 complaining about one speaker dominating the process: "But we are three people", p. 232), alternated between conversation and periods of silence, and little to no metacommunicative comments. This resulted in less interactive and less coordinated discussions, limiting opportunities for shared reasoning and collective construction of knowledge.

https://doi.org/10.48088/ejg.j.mei.17.1.300.319

Furthermore, pronounced differences were found regarding the logic of argumentation. High-performing groups demonstrated more advanced argumentation and contingent thought structures. Often, they employed if-then-logic, such as group 8: “If the terrorism is stopped first, then they actually wouldn’t need to - well, the people behind it would basically lose their purpose anyway. Because if they don’t need weapons anymore, since it’s no longer happening, then they wouldn’t need them at all.”, p. 249. In addition, they used subjunctives, juxtapositions, and hypothetical questions – the latter being used an average of approximately 4 times. These linguistic markers suggest more advanced reasoning, as they allow students to explore alternative scenarios, conditional relationships and complex causal structures. In low-performing groups, such contingent thought structures were almost completely absent. For example, hypothetical questions were only asked once per discussion on average.

Generally, key differences between groups emerged both in the structure and quality of collaboration. While high-performing groups engaged in interactive, responsive and reasoning-oriented discussions, low-performing groups showed more fragmented and less coordinated interaction patterns. These findings suggest that collaborative dynamics were closely intertwined with systems modelling processes.

5.5. Metacognitive Strategies

The usage of metacognitive strategies differed markedly between groups. On average, high-performing groups made approximately seven metacognitive comments during the construction of the concept maps (e.g., group 8: “Should we pull this down a bit, because the thing with the power, it’s also connected to them not having money”, p. 256), while weaker groups only made an average of around two metacognitive comments. These differences suggest that high-performing groups engaged more frequently in monitoring and regulating their own problem-solving processes.

Aside from these quantitative differences, the groups also showed qualitative differences regarding their mapping strategy during the construction process. Low-performing groups went straight into map construction without a pre-discussed plan: immediately starting to glue down information cards and drawing connections led to them having to make more changes later on (e.g. group 12: “Oh no, you drew the wrong arrows”, p. 259). This suggests a more reactive approach in which actions were less frequently preceded by explicit planning or reflection. In contrast, high-performing groups planned the layout of their concept map in advance: “Then we could use ‘Somalia as a failed state’ as the main category, and then we could connect it with the terrorist groups and the—” [gestures toward the positions of card K12 and the related arrows] - R: “Like a tree diagram?” (Group 8, p. 257). Furthermore, they also created semantic categories to organise the information cards, such as “fishery” (group 8), before gluing them down. These behaviours illustrate anticipatory planning and strategic organisation and were coded as metacognitive regulation.

Another key difference was observed concerning reactive vs. proactive modelling strategies. High-performing groups demonstrated proactive modelling approaches, with approximately five concepts actively formed per group. When new concepts emerged during the discussion in the mapping process, these were added to their concept map, suggesting that high-performing groups actively monitor their evolving understanding and flexibly adapt their models during the process. In contrast, low-performing groups primarily showed reproductive strategies by mainly focusing on pre-existing information cards and not going beyond the information given or inferring new concepts from it. This was reflected in a comparatively low number of newly formed concepts, which averaged approximately one per low-performing group. Within the present dataset, such behaviour was associated with fewer instances of metacognitive regulation, as groups largely relied on the given information rather than actively extending or revising their models.

In summary, clear differences between high- and low-performing groups emerge in the extent and quality of metacognitive engagement. While high-performing groups continuously monitored, planned, and adapted their modelling processes, low-performing groups tended to proceed in a more reactive and less reflective manner. Within the present dataset, metacognitive regulation consistently co-occurred with strategic information processing and collaborative reasoning during concept map construction.

5.6. Interrelations between dimensions

Although the five dimensions are presented for analytical clarity, the data indicate that they operated as closely interrelated cognitive, metacognitive, and collaborative processes rather than as isolated activities. Strategic information organisation was frequently accompanied by metacognitive regulation (i.e., group 6), as students explicitly planned how to approach the task (i.e., by explicitly discussing how to organise the information cards) and closely monitored whether their strategy remained effective. This suggests that metacognitive regulation may have been associated with more strategic information processing, as groups coordinated and adapted their approach throughout the task. Similarly, collaborative dynamics appeared to be closely linked to concept and connection formation. Groups characterised by balanced participation, responsive turn-taking, and elaborate dialogue also tended to develop more complex explanations and integrate different systemic features. In these groups, collaborative interaction co-occurred with more sophisticated systems models and with reasoning that was more explicit and open to negotiation. Altogether, these findings suggest that systems modelling competence emerges from the integration of multiple process dimensions rather than discrete skills. Thus, coordination of information processing, reasoning, collaboration and metacognitive regulation was a key distinguishing feature of high- and low-performing groups. Accordingly, higher-level systems modelling appears to depend on the coordination of these dimensions rather than on the isolated presence of individual skills. In total, these five dimensions answer RQ 2 by representing the key differential categories through which systems modelling processes can be analytically distinguished between high- and low-performing groups. However, the qualitative design does not allow conclusions about causal or hierarchical relationships between the dimensions. While metacognitive regulation appeared particularly central because it frequently co-occurred with other dimensions, the present study cannot determine whether any single dimension is dominant in the development of systems modelling competence. This question warrants further investigation using comparative and longitudinal research designs.

6. Discussion

In this video-based study, the process of systemic thinking during the completion of a geographical mystery task was investigated using qualitative content analysis. The results show that systems thinking manifests not only in the product, but also in the process of task completion and differs substantially between high- and low-performing groups. While differences between high- and low-performing groups may appear expected, the contribution of this study lies in revealing how these differences manifest in observable process patterns.

Rather than merely confirming performance differences, the study provides insight into the mechanisms through which systems thinking develops in interaction: High-performing groups acted qualitatively differently throughout the entire concept mapping process. Differences from lower-performing groups became evident in the structure of their discussions, conceptualisation of the maps, information processing, the depth of argumentation, and the reflexivity of their approach. As such, systems thinking unfolds through a variety of social, dialogic, cognitive and metacognitive processes. This finding aligns with the existing body of research conceptualising systems modelling as a processual and practice-based activity rather than a purely cognitive outcome (cf. Ammoneit et al., 2024; Upmeier zu Belzen et al., 2019; Schwarz et al. 2009). Systems modelling is socially distributed and linguistically mediated, as it unfolds through interaction, discourse and shared meaning-making. While this study focuses on modelling processes, the identified patterns align with previous research showing that higher-quality processes are associated with more complex and integrated concept maps (cf. Benninghaus et al., 2019a), suggesting that specific process patterns (e.g., metacognitive regulation, dialogic reasoning, and explicit relation building) are closely linked to the development of more complex and integrated modelling outcomes.

Key differences between groups already became apparent in their approaches to information processing and problem analysis, which can be seen as an indicator of cognitive control. High-performing groups approached the mystery task by systematic reading, sorting, and categorising of the information cards – a structured approach that can be interpreted as an expression of metacognitive regulation (cf. Flavell, 1979). This ability to self-regulate thinking processes is a core aspect of systems competence that has also been embedded in the systems thinking hierarchy (STH) model by Ben-Zvi Assaraf & Orion (2005). Current modelling research also emphasises that an exclusive focus on the end product proves to be insufficient. Instead, metacognitive regulation plays a central role in successful modelling processes, as it enables learners to purposefully plan, monitor, and flexibly adapt their modelling strategies within the context of complex systems (cf. Ammoneit et al., 2024; Göhner et al., 2022). These results also imply a strategic problem awareness that makes it possible to recognise and question linear patterns and replace them with more complex modelling approaches (cf. Rempfler & Uphues, 2011).

Further notable differences emerged regarding the elaboration of concepts and concept formation processes. What was particularly striking about high-performing groups is their ability to introduce new concepts independently (i.e., "governance vacuum", "transnational interests"). These extensions of concepts, meaning going beyond the information given, indicate a deep understanding of systemic interrelationships and the ability to apply and transfer knowledge (cf. Mehren et al., 2018). In contrast, low-performing groups mainly worked reproductively, i.e. operating with the given information but not introducing new and/or related concepts. In education theory, the formation of concepts is a central mechanism for the development of understanding (cf. Vygotsky & Kozulin, 1986). The ability to grasp complex phenomena through precise language is, thus, not merely a side effect, but an integral part of systems thinking. This aligns with findings that concept mapping supports the integration of multiple system components and deeper conceptual understanding when learners actively construct and extend their knowledge structures (cf. Lenski et al., 2022).

In addition, the introduction of new concepts during the modelling process indicates that high-performing groups develop their concept maps dynamically, which was also reflected in their adaptation of the concept map layout during the process – in contrast to low-performing groups, whose layouts primarily remained static. Dynamic modelling, however, indicates a deeper understanding of open problem solving. This flexibility in thought and execution is a crucial characteristic of system competence because complex systems are characterised by unpredictability, emergence and openness (cf. Meadows, 2008). Static and fixed thinking, consequently, inhibits the understanding of complex systems, which was clearly demonstrated in the map progressions of the low-performing groups.

Different levels of linguistic precision were evident not only in concept formation, but also relationship formation. High-performing systems thinkers used precise descriptors to define connections between concepts, such as "causes", "prevents", or "reinforces", showing that they have a differentiated repertoire of relation types and use them in a targeted, precise manner to model complex systems. This ability is central from a systems theory perspective: complex systems are not characterised by simple, linear causalities, but rather by non-linear, bidirectional, and recursive interactions (cf. Sterman, 2002). To model these complex relationships and accompanying dynamics requires linguistic abilities to articulate qualitative relations between concepts in a differentiated manner. In this sense, systems modelling can be understood as a linguistically mediated practice, in which meaning is constructed and negotiated through language (cf. Schuler et al., 2018). This was largely absent in the low-performing groups, who mostly described unspecific relations such as "X makes Y" or "X is Y", indicating a lack of systemic structure not only in the relations, but also their linguistic descriptors and therefore pointing towards a lack of semantic relationship differentiation and deficits in causal-analytical thinking.

Further differences emerged on the dimension of handling advanced systemic features, in particular with regard to handling uncertainty and complexity. High-performing groups showed hypothetical thinking, such as openly speculating about alternative developments, asked counterfactual questions ("What if…?") and discussed from multiple points of view (local vs. global, EU vs. Somalia). These thought processes correspond to what Rieckmann (2012) describes as future thinking, i.e. the ability to anticipate future effects of current decisions. Lower-performing groups, on the other hand, tended to seek conclusive explanations without allowing for alternatives or discussing long-term consequences, indicating a way of thinking that remains static and one-dimensional. This supports findings that dealing with complexity, uncertainty, and multiple perspectives is a key characteristic of systems competence, particularly in the context of sustainability-related issues (Demssie et al., 2023; Nielsen & Nielsen, 2021). The findings of this study also highlight the importance of pedagogies for complexity, as they show that systems modelling emerges through dialogic, collaborative, and reflective processes.

Another key finding is that systems modelling in this study emerged as a collaborative process of co-construction rather than solely an individual activity. High-performing groups showed dialogic and collaborative communication styles, while low-performing groups' discussions are mainly unbalanced and monologic. Successful groups consistently show high quality of communication about the modelling process, manifesting in frequent and cooperative turn-taking, referencing each other, metacognitive statements such as what-if-questions, and negotiating alternative approaches and perspectives. Such forms of discussions are described as exploratory discourse in the literature (Mercer, 2002), referring to dialogic forms of collaborative thinking in which arguments are weighed, alternatives are considered, and uncertainties are made transparent. Exploratory discourse promotes deep learning and, as such, is particularly valuable in complex systemic learning situations (cf. Mercer & Howe, 2012). This finding reinforces recent research highlighting the central role of interaction and discourse in the development of systems competence, as knowledge is co-constructed through language and collaborative reasoning (Schuler et al., 2018).

In contrast, low-performing groups mainly engaged in cumulative talk with monologic dominance of individual speakers – a type of discourse that is significantly less suitable for fostering or reflecting on systemic modelling processes (cf. Mercer & Howe, 2012). Our findings therefore suggest that the collaborative format of mystery tasks provides opportunities for students to externalise, negotiate, and refine systems models through dialogue. However, as this study did not include an individual comparison condition, no conclusions can be drawn about whether collaborative mystery tasks are more beneficial than individual systems modelling.

Additionally, this key finding can also explicitly be tied back to systems thinking research, aligning with results by Karkdijk et al. (2019) who showed that on-task discussion is the strongest predictor of higher-level outcomes according to the SOLO-taxonomy (cf. Biggs & Collis, 1982) – as is the case for high-performing groups. However, the results of this study partly contradict Hempowicz' claim that "the mystery method hardly stimulates discursive engagement with the content" (2021, p. 203). Clearly, high-performing groups showed a high degree of dialogic and collaborative communication while negotiating the systemic modelling of the content. One possible explanation for this may lie in the different mystery topics used: while Hempowicz' study used the case example of "rose farming at lake Naivasha (Kenia)" (p. 65), this study used the example of "piracy in Somalia". Also given the empirical finding described in the literature that systems competence is domain-specific (cf. Mehren et al., 2018), the question arises to what extent findings on geographical systems competence can be transferred across different systemic contents. Further studies should examine this using different types of systems (i.e., human system, environment system, human-environment system) to investigate transferability and discursive engagement during mystery tasks.

6.1. Limitations

Despite the multi-layered insights into the development of systems thinking during the collaborative construction of concept maps, this study has several limitations regarding its scope, focus and methodological approach. First, the sample size at n=30 students is, as typical for qualitative research, relatively small. Together with recruitment in only one German federal state, this limits generalisability of the findings. However, the quality control measures in place (see section 4.3.2) ensure the trustworthiness of this study. In addition, the differentiation between high- and low-performing groups is based on a pretest of systems competence. While this enables a contrastive analysis of process patterns, the operationalisation of systems competence through a single assessment remains limited and cannot fully capture the complexity and situational variability of students' abilities. At the same time, the use of homogeneous groups (high- vs. low-performing) represents a specific analytical design choice in accordance with the research interest. Future research should therefore explore how systems modelling processes develop in heterogeneous group constellations. Tying in with this, the categorisation of students into high- and low-performing groups may risk implying stable ability differences. In this study, however, these categories are understood as analytical constructs describing situationally emerging process patterns rather than fixed individual competencies.

Secondly, the sample strategy does not allow the isolation of individual contributions of students. From a constructivist perspective, however, learning is understood as a collaborative process, and mystery tasks are inherently designed for collaborative work. Furthermore, the group setup mimics classroom practice, in which students mainly work on similar tasks in groups, bringing the findings closer to classroom practice and enhancing their relevance for geography teaching. In addition, the study does not include students' subjective perspectives on their own learning processes. Insights into how students perceive and reflect on their learning could provide an additional layer of understanding and should be considered in future research.

Thirdly, observer effects due to the presence of cameras cannot be fully ruled out. However, this was accounted for in the methodological setup of this video-based study: Students were made familiar with the camera and recording equipment in advance of filming, following best practices to reduce observer reactivity observed in video-based classroom studies (cf. Liang, 2015), thereby minimising potential observer effects.

Fourthly, it must be taken into account that the data collection took place a decade ago. This could limit the significance of the results regarding current strategy use patterns, especially as research has progressed in the field of systems thinking over the last ten years (cf. Pivarníková, 2025). It cannot be ruled out that current learners already possess a pronounced understanding of systems due to increased school-based interventions. Nevertheless, systemic thinking is not an intuitive cognitive ability and must be learned through interventions (e.g., Tripto et al., 2017; Ben-Zvi Assaraf & Orion, 2005). The present findings are not intended to reflect the current state of development but rather to analyse the strategic patterns of high- and low-performing groups, which remain significant for both research and practice.

https://doi.org/10.48088/ejg.j.mei.17.1.300.319

Furthermore, while extensive efforts were made to ensure transparency of the qualitative analysis, including detailed coding procedures, interrater agreement, and peer debriefing, qualitative video analysis inherently involves interpretative decisions. Therefore, full reproducibility in a strict quantitative sense cannot be achieved. Due to the necessary reduction and aggregation of qualitative data into higher-order categories, not all observed strategies and interaction patterns could be represented in the final analysis. While care was taken to capture the most relevant and contrasting patterns, the results inevitably reflect a selection and abstraction of the full data corpus. In this context, the use of frequency-based indicators (e.g., the number of metacognitive statements) should be interpreted with caution. These indicators serve as heuristic devices to illustrate tendencies within qualitative data rather than as statistically robust measures. Moreover, the study focuses exclusively on process-oriented data and does not systematically assess learning outcomes or the quality of the final concept maps. Therefore, no direct conclusions can be drawn regarding the relationship between observed processes and learning success. Finally, the findings are based on a specific instructional format (geographical mystery task), and their transferability to other modelling contexts should therefore be considered with caution.

Despite these limitations, the strength of this design lies in capturing systems thinking as it occurs, offering insights into the collaborative strategies through which students model concept maps. By focusing on process-oriented data, the study contributes to a more nuanced understanding of how systems modelling competence develops through interaction, complementing existing product-oriented research approaches.

7. Conclusions and Outlook

7.1. Implications for Teaching and Professionalisation

As the study deliberately examined homogenous high- and low-performing groups, the following implications are derived from these specific group constellations, with their applicability to heterogenous classrooms remaining to be investigated empirically. Implications can be grouped into three interrelated areas: 1) systemic lesson design, 2) process-oriented diagnostics, and 3) teacher training and professionalisation.

- *Systemic Lesson Designs:* The findings suggest that effective systems modelling depends not only on the quality of the final concept map but also on the collaborative processes through which it is constructed. Lesson designs could therefore explicitly scaffold these processes rather than focusing exclusively on the final product. For example, collaborative learning settings can be supported through targeted role allocation (i.e., "system coach", "moderator", "critical friend", "note-taker" …). This can be enhanced further through targeted metacognitive impulses (i.e. "How confident are you about [this connection/ placement of cards/ …], "What types of interactions could exist between them?"). Lower-performing groups, in particular, may benefit from such guidance, as they tended to show fewer spontaneous metacognitive strategies. In instructional contexts with heterogeneous group compositions, there is the possibility that lower-achieving learners can observe metacognitive strategies used by higher-achieving peers and adapt them in subsequent learning situations; however, this possibility was not examined in the present study. Visualisation tools can further support students' modelling, for example via mapping layouts or partly pre-structured maps (cf. Cox et al., 2022). While this study used classic pen-and-paper designs, dynamic concept mapping software or apps allow real-time visualisation and facilitate easy revisions of the concept map. Finally, lesson designs should explicitly allocate time for hypothetical, critical, and divergent thinking, for example through targeted reflection tasks, and integrate structured reflection and feedback phases into classroom practice.
- *Diagnostics:* Furthermore, this study underscores the value of process diagnostics in teaching geography, since the path to a concept map is just as revealing as the outcome. Individualised feedback and support are necessary to promote effective systems thinking strategies on cognitive, metacognitive, linguistic, and social levels. For diagnostic purposes, the group composition is of secondary importance, as the learners' level of knowledge is evaluated individually. Nevertheless, the results of this study provide meaningful indications for identifying individual levels of learning as well as for determining developmental potential. To make process diagnostics more accessible in everyday teaching, these findings should be translated into an applied diagnostic tool, which could assess several dimensions of systems thinking: (1) cognitive practices, including diversity and quality of concept selection and relation formation; (2) advanced systemic features; (3) communicative practices, distinguishing explanatory-discursive versus monological approaches; and (4) metacognitive strategies, distinguishing reactive from proactive strategies. These categories can be used to provide formative

feedback or to document development processes. Existing diagnostic tools provide valuable approaches for assessing the outcomes of systems thinking in geography education (cf. Ben-Zvi Assaraf & Orion, 2005; Mehren et al., 2018) and ESD (cf. Benninghaus et al., 2019b), but do not focus on the underlying processes. By developing a process-oriented diagnostic tool for systemic thinking, teachers can observe (i.e., through conversation transcripts or in vivo), evaluate (i.e., through criteria grids), provide feedback (i.e. regarding the depth of argumentation, use of technical terms…), and promote (i.e., through questioning techniques or support cards) systems thinking processes of their students.

- *Teacher training and professionalisation:* Providing adequate diagnostic feedback and individual support requires adequate teacher expertise: they must be able to recognise, evaluate, and scaffold systems thinking in real time. Recent studies indicate that individual feedback in geography classrooms is often limited (cf. Bienert et al., 2024), with no evidence that systems competence is an exception. Developing professional training modules and workshops (i.e., video-based classroom analyses, collaborative coding of student interactions while working on a mystery task, …) focused on process-oriented diagnostics and scaffolding strategies (i.e., habits of a systems thinker, Waters Center for Systems Thinking, 2020) is therefore essential to support teachers in promoting systems thinking effectively.

7.2. Implications for Research

While this qualitative study is a step towards process-oriented research in geographical systems thinking, it also reveals methodological and conceptual pathways for future research.

First, long-term process observations are needed to reveal changes in systemic thinking learning processes. While this study focused on short-term development of systems modelling competence, it would be interesting to investigate how the systemic strategies, metacognitive routines and collaborative work of students develop over longer periods of time, especially between high- and low-performing students. Possibly, this could be combined with intervention studies to identify which interventions promote positive changes in these strategies. Second, future research should employ intervention designs with pre-post intervention designs. By specifically developing methods that support systems thinking (i.e., Socratic dialogue, digital concept mapping, structured task-based group discussions), such designs could evaluate the effectiveness of these methods to further optimise them for classroom practice. Third, linking qualitative and quantitative process data seems a promising approach. By triangulating process observations with quantitative evaluations, such as network analysis (cf. Hempowicz, 2021), typical patterns of "effective systems thinking" can be identified even more reliably, making it possible to further develop diagnostic tools for either self-assessment or formative assessment. On the whole, the present, process-based study provide insights into the modelling dimension of systems competence (cf. Mehren et al., 2018) by highlighting differences in processes between high- and low-performing students. However, this represents only one dimension of systems competence. As empirical research shows, the second dimension of systems competence – system-adequate intention to act and prognoses – does not correlate with the first dimension (Mehren et al., 2018). Thus, even high-performing students in the present study might be able to model systems in an advanced way, while their system-adequate prognoses might not correspond accordingly. Consequently, a central friction point arises: How do high- and low-performing students form systemic prognoses and intentions to act? This question will explored in future research.

Funding: This research received no external funding.

Ethics Declaration: All participating students and their parents provided written informed consent to take part in the study and to be video-recorded for research purposes. Participation was entirely voluntary, and all data were anonymised prior to analysis and publication. The study was carried out in accordance with institutional ethical standards of the General Data Protection Regulation (DSGVO) of Germany.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix: Coding Scheme

Dimension (theoretically derived)	Definition	Subcategory (empirically derived)	Definition	Coding Rule	Example
Information Processing	Selection, structuring, and organisation of information cards	Structured Reading	Systematic engagement with all information cards before modelling	Code when group explicitly aims to gain overview	"Should I read all cards first?"
		Selective Reading	Fragmented or ad hoc reading of individual cards	Code when cards are read without overarching plan	"Let's just take this one"
		Categorisation	Grouping information into thematic clusters	Code when cards are explicitly grouped	"These belong together – economy"
		Highlighting Relevance	Marking or discussing importance of information	Code when relevance is explicitly evaluated	"This card is important"
		Early Hypothesis formation	Initial explanations before full analysis	Code when early causal assumptions are made	"Maybe it's because of poverty"
Concept and Category Formation	Development of concepts and explicit relationships between system components	Reproductive concept use	Using only given concepts	Code when no new abstraction occurs	"Just use what's written here"
		Concept extension	Introducing new concepts beyond given material	Code when new systemic terms emerge	"This is like a governance problem"
		Explicit relation naming	Defining type of connection	Code when relation type is verbalised	"This reinforces that"
		Implicit linking	Connections without specification	Code when links are made without explanation	"That goes here"
		Negotiated connections	Discussion and revision of relations	Code when connections are debated	"Or maybe it connects differently?"
Dealing with Advanced Systemic Features	Engagement with complexity, feedback loops, uncertainty, multiperspectivity, and temporal dynamics	Linear reasoning	Simple cause–effect chains	Code when only one-directional reasoning appears	"A leads to B"
		Feedback thinking	Recognition of circular relationships	Code when loops are identified	"It goes in a circle"
		Multiperspectivity	Considering different viewpoints	Code when perspectives are contrasted	"For the EU it's different than for locals"
		Hypothetical reasoning	Use of "what if" scenarios	Code when alternative scenarios are discussed	"What if piracy stopped?"
		Dealing with uncertainty	Expressing doubt or alternatives	Code when uncertainty is verbalised	"Maybe… or maybe not"
Collaborative Dynamics	Interaction patterns, participation, discourse structure, and co-construction of meaning	Balanced participation	All group members actively contribute	Code when all students speak regularly	All three contribute evenly
		Dominance	One student dominates interaction	Code when one speaker controls discussion	"Let me do it"
		Turn-taking	Interactive exchange of ideas	Code when ideas are taken up and extended	"Yes, and then…"
		Referential talk	Referring to others' contributions	Code when previous statements are used	"Like she said…"
		Argumentative reasoning	Justifying and evaluating ideas	Code when arguments are elaborated	"Because otherwise…"
		Planning	Explicit organisation of approach	Code when strategy is discussed beforehand	"Let's plan first"

https://doi.org/10.48088/ejg.j.mei.17.1.300.319

Metacognitive Strategies	Planning, monitoring, evaluating, and adapting modelling processes	Monitoring	Checking progress or understanding	Code when group evaluates own process	"Does this make sense?"
		Regulation	Adjusting strategy during task	Code when approach is changed	"Let's move this"
		Error correction	Identifying and fixing mistakes	Code when errors are noticed	"That arrow is wrong"
		Proactive modelling	Anticipatory strategy use	Code when actions are planned ahead	"We should structure first"

References

Ammoneit, R., Göhner, M., Bielik, T., & Krell, M. (2024). Why most definitions of modeling competence in science education fall short: Analyzing the relevance of volition for modeling. *Science Education, 108*(2), 443–466. https://doi.org/10.1002/sce.21841

Aufschnaiter, C., von, Münster, C., & Beretz, A. K. (2018). Zielgerichtet und differenziert diagnostizieren. *MNU Journal*, 71(6), 382–387.

Benninghaus, J. C., Mühling, A., Kremer, K., & Sprenger, S. (2019a). The mystery method reconsidered - A tool for assessing systems thinking in education for sustainable development. *Education Sciences, 9*(4), 260. https://doi.org/10.3390/educsci9040260

Benninghaus, J. C., Mühling, A., Kremer, K., & Sprenger, S. (2019b). Complexity in education for sustainable consumption - An educational data mining approach using mysteries. *Sustainability, 11*(3), 722. https://doi.org/10.3390/su11030722

Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, 42(5), 518–560. https://doi.org/10.1002/tea.20061

Ben-Zvi Assaraf, O., Dodick, J., & Tripto, J. (2013). High School Students' Understanding of the Human Body System. *Research in Science Education,* 43, 33–56. https://doi.org/10.1007/s11165-011-9245-2

Bienert, N., Scholten, N., & Mehren, R. (2024). Unterrichtsqualität in der Geographie – Eine fachspezifische Sekundäranalyse der sächsischen Schulevaluation. Zeitschrift für Geographiedidaktik (ZGD), 51(3), 123–142. https://doi.org/10.60511/51184

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

Cañas, A. J., Reiska, P., & Shvaikovsky, O. (2023). Improving learning and understanding through concept mapping. *Knowledge Management & E-Learning*, 15(3), 369–380. https://doi.org/10.34105/j.kmel.2023.15.021

Chang, C.-C., Liu, G.-Y., Chen, K.-J., Huang, C.-H., Lai, Y-M., & Yeh, T.-K. (2017). The Effects of a Collaborative Computer-based Concept Mapping Strategy on Geographic Science Performance in Junior High School Students. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(8), 5049 – 5060. https://doi.org/10.12973/eurasia.2017.00981a

Cox, M., Elen, J., & Steegen, A. (2022). A Test to Measure students' Systems Thinking Abilities in Geography. *European Journal of Geography, 9*(1), 105-120. https://www.eurogeojournal.eu/index.php/egj/article/view/102

Demssie, Y. N., Biemans, H. J. A., Wesselink, R., & Mulder, M. (2023). Fostering students' systems thinking competence for sustainability by using multiple real-world learning approaches. *Environmental Education Research*, 29 (2), 261-286. https://doi.org/10.1080/13504622.2022.2141692

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3–53. https://doi.org/10.1080/10508400903452884

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Göhner, M. F., Bielik, T., & Krell, M. (2022). Investigating the dimensions of modeling competence among preservice science teachers: Meta-modeling knowledge, modeling practice, and modeling product. *Journal of Research in Science Teaching, 59(8)*, 1354-1387. https://doi.org/10.1002/tea.21759

Hempowicz, J. (2021). *Systemorganisationskompetenz im Geographieunterricht: Videobasierte Fallanalysen von Schü-ler*innen im Rahmen der Mystery-Methode* (Vol. 75). BoD–Books on Demand.

Karkdijk, J. (2021). *Mysteries to support geographical relational thinking in secondary education* [Doctoral dissertation, Vrije Universiteit Amsterdam]. Quaeris Media BV.

Karkdijk, J., van der Schee, J., & Admiraal, W. (2021). Strategies used by small student groups to understand a geograph-ical mystery. *Journal of Research and Didactics in Geography,* 1, 5–21.

Karkdijk, J., Admiraal, W., & van der Schee, J. (2019). Small-group work and relational thinking in geographical mysteries. *Review of International Geographical Education Online, 9*(2), 402–425. https://doi.org/10.33403/rigeo.588661

Khajeloo, M., & Siegel, M. A. (2022). Concept map as a tool to assess and enhance students' system thinking skills. *Instructional Science*, 50, 571-597. https://doi.org/10.1007/s11251-022-09586-5

Kleemann, F., Krähnke, U., & Matuschek, I. (2009). *Interpretative Sozialforschung: Eine praxisorientierte Einführung*. VS Verlag für Sozialwissenschaften.

Kuckartz, U., & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung: Grundla-gen-texte Methoden* (5th ed.)*.* Juventa Verlag.

Kwon, S. Y., & Cifuentes, L. (2009). The comparative effect of individually-constructed vs. collaboratively-constructed computer-based concept maps. *Computer & Education*, 52 (2), 365 – 375. https://doi.org/10.1016/j.compedu.2008.09.012

Lamb, C. M. T. (2009). *Collaborative systems thinking: An exploration of the mechanisms enabling systems thinking*. [Doctoral dissertation, Massachusetts Institute of Technology].

Leat, D. (1998). *Thinking through geography*. Chris Kington Publishing.

Leat, D., & Nichols, A. (1999). *Mysteries make you think*. Chris Kington Publishing.

Leat, D., & Nichols, A. (2000). Brains on the Table: Diagnostic and formative assessment through observation. *Assess-ment in Education: Principles*, *Policy & Practice* 7(1), 103–121. https://doi.org/10.1080/713613327

Lenski, S., Elsner, S., & Großschedl, J. (2022). Comparing construction and study of concept maps – an intervention study of learning outcome, self-evaluation and enjoyment through training and learning. *Frontiers in Education, 7,* 892312. https://doi.org/10.3389/feduc.2022.892312

Li, R., & Li, G. (2023). Exploring Lower-Secondary School Students' Systems Thinking Performance in Ecological Issue. *Journal of Baltic Science Education,* 22(5), 865–880. https://doi.org/10.33225/jbse/23.22.865

Liang, J. (2015). Live video classroom observation: an effective approach to reducing reactivity in collecting observa-tional information for teacher professional development. *Journal of Education for Teaching, 41*(3), 235–253. https://doi.org/10.1080/02607476.2015.1045314

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz.

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.

Mehren, R., Rempfler, A., Buchholz, J., Hartig, J., & Ulrich-Riedhammer, E. M. (2018). System competence modelling: Theoretical foundation and empirical validation of a model involving natural, social and human–environment sys-tems. *Journal of Research in Science Teaching,* 55(5), 685 - 711. https://doi.org/10.1002/tea.21436

Meister, J. (2020). Systemisches Denken bei Schülerinnen und Schülern unterschiedlicher Kompetenzstufen. Eine qua-litative Prozess- und Produktanalyse am Beispiel der Bearbeitung eines geographischen Mysterys [Doctoral dis-sertation, Justus-Liebig-Universität Gießen]. JLUpub. http://dx.doi.org/10.22029/jlupub-9837

Mercer, N. (2002). *Words and minds: How we use language to think together*. Routledge. https://doi.org/10.4324/9780203464984

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction, 1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001

Nielsen, S. S., & Nielsen, J. A. (2021). A competence-oriented approach to models and modelling in lower secondary science education: Practices and rationales among Danish teachers. *Research in Science Education*, *51*(Suppl 2), 565-593. https://doi.org/10.1007/s11165-019-09900-1

Pivarníková, V. (2025). Utilisation of Concept Maps in Geography Education Research: A Systematic Review. *European Journal of Geography* 16(2), 169-184. https://doi.org/10.48088/ejg.v.piv.16.2.169.183.

Rempfler, A., & Uphues, R. (2011). Systemkompetenz im Geographieunterricht – Die Entwicklung eines Kompetenzmodells. In *Geographische Bildung: Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 36–48).
Rieckmann, M. (2012). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures, 44*(2), 127–135. https://doi.org/10.1016/j.futures.2011.09.005
Roczen, N., Fischer, F., Fögele, J., Hartig, J., & Mehren, R. (2021). Measuring system competence in education for sustainable development. *Sustainability, 13*(9), 4932. https://doi.org/10.3390/su13094932
Schuler, S., Fanta, D., Rosenkraenzer, F., & Riess, W. (2018). Systems thinking within the scope of education for sustainable development (ESD) – a heuristic competence model as a basis for (science) teacher education. *Journal of Geography in Higher Education, 42*(2), 192-204. https://doi.org/10.1080/03098265.2017.1339264
Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632–654. https://doi.org/10.1002/tea.20311
Sterman, J. D. (2002). System dynamics: Systems thinking and modeling for a complex world. *[Massachusetts Institute of Technology. Engineering Systems Division].* https://dspace.mit.edu/handle/1721.1/102741
Tripto, J., Ben-Zvi Assaraf, O., Snapir, Z., & Amit, M. (2017). How is the body's systemic nature manifested amongst high school biology students. *Instructional Science*, 45, 73-98. https://doi.org/10.1007/s11251-016-9390-0
Tuma, R., Schnettler, B., & Knoblauch, H. (2013). *Videographie: Einführung in die interpretative Videoanalyse sozialer Situationen*. Springer VS.
Upmeier zu Belzen, A., van Driel, J., & Krüger, D. (2019). Introducing a Framework for Modeling Competence. In Towards a Competence-Based View on Models and Modeling. In A. Upmeier zu Belzen, D. Krüger, & J. van Driel (Eds.): *Towards a Competence-Based View on Models and Modeling in Science Education. Models and Modeling in Science Education* (pp. 3-19). Springer . https://doi.org/10.1007/978-3-030-30255-9_1
Vygotsky, L. S. & Kozulin, A. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
Waters Center for Systems Thinking (2020). Habits of a systems thinker, Thinking Tools Studio: Habits of a Systems Thinker. Available at: https://thinkingtoolsstudio.waterscenterst.org/cards (Accessed: 06 July 2026).
Yoon, S. A., & Hmelo-Silver, C. (2017). Introduction to special issue: models and tools for systems learning and instruction. *Instructional Science, 45*, 1–4. https://doi.org/10.1007/s11251-017-9404-6
Zohar, A., & Dori, Y. J. (Eds.). (2012). *Metacognition in science education: Trends in current research*. Springer. https://doi.org/10.1007/978-94-007-2132-6

https://doi.org/10.48088/ejg.j.mei.17.1.300.319