

---

# The *Art Nouveau Path*: Requirements Engineering and Traceability for City-Scale In-the-Wild Mobile Augmented Reality Games Learning Services

---

[João Ferreira-Santos](#)<sup>\*</sup> and [Lúcia Pombo](#)<sup>\*</sup>

Posted Date: 2 March 2026

doi: 10.20944/preprints202603.0012.v1

Keywords: software engineering; requirements engineering; traceability; mobile computing; human-computer interaction (HCI); augmented reality; reproducibility; heritage-based learning; educational software



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# The Art Nouveau Path: Requirements Engineering and Traceability for City-Scale In-the-Wild Mobile Augmented Reality Games Learning Services

João Ferreira-Santos \* and Lúcia Pombo \*

CIDTFF – Research Centre on Didactics and Technology in Education of Trainers, Department of Education and Psychology, University of Aveiro, 3810-193 Aveiro, Portugal

\* Correspondence: joaomrsantos@ua.pt (J.F.-S.); lpombo@ua.pt (L.P.)

## Abstract

City-scale, in-the-wild Augmented Reality (AR) learning paths must remain operable under Bring Your Own Device (BYOD) heterogeneity, outdoor tracking degradation, public-space safety, and interruption recovery. This study conceptualizes the *Art Nouveau Path* as an AR learning service and makes a theoretical contribution by proposing a Determinant-driven Requirements traceability model that treats implementation Determinants as Requirements signals and links them to testable Requirements, transfer Artefacts, and evidence anchors for replication. Methods combined 8 Points of Interest (POIs) and 36 tasks profiling, group-session logs (118 sessions), and teacher-facing records from a validation workshop (T1-VAL, N=30) and in situ observation (T2-OBS, N=24). Teachers open-text fields were segmented into meaning units and coded with an eight-Determinant taxonomy, with intercoder reliability assessed on a stratified subset (Krippendorff's alpha = 0.83). Logs and a post-path student questionnaire (S2-POST, N=439) bounded enactment feasibility and data integrity, without learning-outcome inference. Dominant determinants concerned onboarding and legibility, marker robustness and recovery, and curriculum framing, alongside safety and fallback constraints. These signals were translated into 18 "shall" Requirements with acceptance criteria and bidirectional trace links to transfer 6 Artefacts. The resulting transfer kit specifies routines, maintenance, incident handling, and fallback procedures to reduce replication fragility across teams.

**Keywords:** software engineering; requirements engineering; traceability; mobile computing; human-computer interaction (HCI); augmented reality; reproducibility; heritage-based learning; educational software

---

## 1. Introduction

Outdoor mobile AR learning paths are increasingly deployed beyond controlled pilots, moving into city-scale, public-space contexts where the system must operate under variable environmental and organizational conditions. In this shift, the technical object is no longer only an application that "runs", but a service that must remain usable, robust, and governable across heterogeneous devices, locations, and supervision regimes. This service framing aligns with established HCI arguments that "in-the-wild" deployments change the epistemic and engineering problem: what matters is sustained functioning in context, including breakdowns, repairs, and the socio-technical conditions under which use is possible [1].

For city-scale AR, operational reliability is repeatedly challenged by failure modes that are amplified outdoors and on-the-move: First, sensor and tracking performance is sensitive to lighting variability, reflections, and occlusions, which affects camera-based recognition and marker-mediated anchoring under glare and weathered surfaces [2]; Second, physical assets used by AR experiences (for example, printed markers or plaques) are exposed to degradation, vandalism, and routine maintenance constraints, shifting "robustness" from a purely software concern to a coupled cyber-

physical requirement; Third, city enactment typically operates under BYOD conditions, introducing device heterogeneity (Operating Systems (OS) versions, camera quality, memory, thermal throttling), inconsistent permission states, and uneven security posture; these issues are well documented as practical risks in mobile device management and enterprise mobile security guidance [3]; Fourth, mobility and safety constraints become first-order design and operational concerns: attention must be shared between navigation, traffic awareness, group coordination, and device interaction; and, fifth, interruptions are normative in public space (connectivity drops, app switching, group pacing, supervision pauses), making recoverability and state restoration part of the core system specification rather than an edge case.

Empirical work on AR in-the-wild also highlights how real-world social conditions and non-user dynamics introduce friction that typical lab-centered assumptions under-specify, reinforcing the need for operationally grounded deployment design [4]. Accordingly, usability and quality must be treated as outcomes of use in context, not only as interface properties. This view is consistent with International Organization for Standardization (ISO)'s definition of usability as effectiveness, efficiency, and satisfaction in a specified context of use, which becomes a service-level requirement when the "object of interest" is a deployable socio-technical system rather than a standalone artifact [5]. The central engineering challenge, therefore, is to specify and package the requirements, trace links, and operational routines that make a city-scale AR learning service repeatable, inspectable, and responsibly maintainable across cohorts and implementers.

A learning intervention can, conceptually, be considered an educational program aimed at altering learner outcomes under certain conditions, assessed by fidelity, feasibility, and effectiveness during a specific implementation period. In this specific context, replication addresses the use of a standardized approach throughout successive cohorts or contexts, with implementation and the realization of predetermined objectives as the focus [6]. This lens is appropriate for effectiveness questions, but it can under-specify the engineering and governance work required when the same activity must be repeatedly enacted across contexts that are inherently unstable, such as public-space, in-the-wild mobile deployments [1,7].

By contrast, a learning service is an operations-bearing socio-technical system that must remain usable, reliable, and governable across repeated enactments, changing cohorts, heterogeneous devices, and evolving environmental conditions. A service boundary therefore includes not only the software artefact and content package, but also the operational routines that stabilize use in context, such as onboarding, supervision, maintenance of physical assets, incident response, and versioned releases [8]. This distinction has two direct implications. First, sustainability becomes a lifecycle property: the system must be maintainable and operable over time under realistic resource constraints, rather than successful in a single delivery cycle [5,9]. Second, transferability requires an auditable "transfer kit" that packages both technical and operational artefacts so that reliability and quality in use can be re-established by non-originating teams, consistent with usability as an outcome of use in context [5].

Despite substantial studies about applied AR and Mobile Learning (ML) deployments, the evidence base often prioritizes experiential and outcome-oriented reporting, while the engineering specification needed for repeatable, city-scale implementation remains comparatively under-articulated. Therefore, scaling a field-deployed system requires requirements that cover not only functional capabilities but also non-functional requirements (NFRs) and operability properties such as robustness, usability, safety, privacy, maintainability, and recoverability. These concerns map directly onto established requirements engineering expectations for lifecycle requirements information items and their management [5] and to quality models used to structure NFR thinking [10]. A second gap concerns auditability: when a deployment succeeds or fails, it should be possible to trace "why" through explicit links between evidence, requirements, and concrete implementation artefacts.

Requirements traceability has long been framed as a two-sided problem: (i) pre-requirements traceability links requirements to their origins; and, (ii) rationale, whereas post-requirements

traceability links requirements to downstream design, implementation, and verification artefacts [11,12].

Recent works confirm that pre-requirements traceability remains underdeveloped relative to its practical importance, with persistent challenges around workload, versioning, and trust in trace links [13,14].

Practitioner-facing evidence also indicates that traceability is often perceived as costly and difficult to maintain across teams and tools, even when its value is acknowledged [15].

This study targets an implementation-specification gap in Mobile Augmented Reality Games (MARG) research by contributing a determinant-driven, transferable requirements engineering and traceability framework that links in-the-wild evidence to testable requirements and operations-ready artefacts for auditable replication. [1,5,12,13,15].

The present paper addresses these gaps by reconceptualizing the *Art Nouveau Path* as a deployable educational software service and by providing an auditable, determinant-driven requirements specification linked to a requirements-to-artefact traceability model. The scope is intentionally implementation-centric: the goal is not to report learning outcomes, but to specify what must hold for a city-scale, in-the-wild AR learning service to be safely and reliably enacted and replicated. Considering this, presented boundaries align with the broader implementation literature's distinction between implementation outcomes (for example, feasibility) and effectiveness outcomes, emphasizing that a system can fail due to implementation breakdown even when the underlying intervention concept is sound [7,16].

The goal is to produce a determinant-driven requirements specification with explicit traceability from evidence to requirements and transferable implementation artefacts, enabling operations-ready replication beyond the originating team. Accordingly, four inspectable Contributions (C1 – C4) are delivered:

**C1.** Determinant-driven requirements specification. An eight-code implementation determinant taxonomy (D1–D8) was operationalized as requirements signals to translate field evidence into deployable constraints.

**C2.** Verifiable requirements catalogue (REQ-1 to REQ-18). Determinant signals were translated into a minimal set of testable “shall” requirements with verification cues aligned with requirements specification guidance [5,42].

**C3.** Evidence-to-requirement-to-artefact traceability. A compact bidirectional trace spine links determinants, requirements, transfer artefacts, and evidence anchors to support auditability and address pre-requirements traceability gaps [12,13,15,55].

**C4.** Operations-ready transfer kit and minimal operations stack. An operations package specifies roles, routines, maintenance, incident response, and BYOD fallback to support replication in applied computing contexts [41,56].

To operationalize this service framing and its audit requirements, three Research Questions (RQs) guide this study:

**RQ1.** Which implementation determinants concentrate in teachers-facing evidence (in this study, T1-VAL and T2-OBS) when a city-scale, in-the-wild mobile AR learning service is enacted?

**RQ2.** How can determinant signals be systematically translated into a coherent Requirements set that specify functional, non-functional, and operational properties needed for safe and repeatable deployment?

**RQ3.** How can evidence-to-requirement-to-artefact traceability be structured to support auditability, replication, and responsible operations, including privacy-aware and BYOD-constrained enactment?

The remainder of the paper is organized as follows: Section 2 reviews related work on in-the-wild mobile AR reliability and recovery, requirements engineering for socio-technical mobile services, traceability as an audit mechanism, and reproducibility-oriented transfer kits. Section 3 presents the system boundary and architecture, evidence streams, Determinant taxonomy, coding and feasibility descriptor methods, and the determinant-to-Requirements derivation and traceability

procedures. Section 4 reports profiling results, feasibility envelopes, determinant concentrations, the derived requirements set, the compact traceability matrix, and the minimal operations stack and transfer kit outputs. Section 5 discusses determinants as operability and NFR drivers, situates the contribution against prior HCI and requirements engineering literature, and consolidates limitations and threats to validity under an explicit implementation-only scope boundary. Section 6 concludes with a summary of contributions, limitations and future paths.

## 2. Background and Related Work

### 2.1. *In-the-Wild Mobile AR: Reliability, Interruptions, Recovery*

City-scale outdoor mobile AR is influenced by several situations, such as environmental variability, social interactions, and mobile usage constraints or device features. In HCI, in-the-wild research has been framed as a shift from evaluating interaction in controlled settings to understanding how systems perform under real breakdown conditions, including repair, adaptation, and context-driven constraints that are not reducible to interface design alone [1,18]. This perspective is directly relevant to outdoor AR services, where failures are often systematic rather than exceptional.

From a technical standpoint, tracking and anchoring remain core fragilities of AR systems. Outdoor illumination variability, glare, occlusion, and viewpoint instability affect both marker-based recognition and broader tracking pipelines, producing failure modes that must be anticipated at requirements level. Recent synthesis work on AR tracking and implementation risks emphasizes that tracking technologies, toolchains, and security considerations interact in complex ways across devices and contexts, reinforcing the need to treat robustness and recoverability as first-order engineering targets rather than *post hoc* fixes [2]. In marker-mediated experiences, physical marker wear, vandalism, and maintenance constraints introduce a cyber-physical coupling: a “tracking defect” can originate in environmental and material conditions, not only software.

Interruptions are a normal condition of use in public-space enactment. Attention alternates between wayfinding and on-screen interaction; sessions pause due to situational safety and regrouping; connectivity fluctuates; devices are handed across peers; and task flow is disrupted by app switching and OS-level notifications. Controlled mobile studies indicate that more cognitively demanding interruptions increase overall completion time and are experienced as more disruptive, motivating design that privileges recovery and resumption rather than assuming continuous task flow [19]. In ML contexts, task resumption or memory cues have been implemented and evaluated as explicit support for returning to the prior learning state; while objective performance effects can be limited, users still report such cues as helpful in interruption-prone environments, reinforcing the view that continuity is a designable property rather than an emergent user skill [20].

In-the-wild AR introduces additional socio-technical friction: non-users, bystanders, and public norms may define whether AR interaction is socially acceptable and whether group activity remains coordinated. A recent in-the-wild AR study highlights awkwardness and social constraints as practical factors affecting sustained use, indicating that operational success depends on more than technical tracking and interface affordances [4]. Consequently, reliability for city-scale AR services should be interpreted as an end-to-end property spanning tracking, recoverability, social operability, and supervision routines.

### 2.2. *Requirements Engineering for Socio-Technical Mobile Services*

Requirements engineering provides the methodological basis for translating stakeholder and operational evidence into verifiable system properties. ISO standard 29148 formalizes Requirements engineering as a life-cycle process and specifies the information items expected for systems and software products, explicitly including services [5]. For city-scale mobile AR services, this framing is important because the system boundary necessarily includes socio-technical operations: briefing, supervision, path pacing, fallback routines, and maintenance of physical assets.

A defensible requirements specification must integrate functional requirements with NFRs and operational constraints. ISO's 25010 structures product quality in terms of characteristics such as reliability, usability, security, maintainability, and portability, supporting systematic NFR elicitation and acceptance criteria definition [10]. Similarly, usability in ISO 9241-11 is defined as an outcome of use in context, which is aligned with the service framing of in-the-wild deployment: usability cannot be assumed stable across contexts because context is part of the specification target [5].

Mobile AR services operating under BYOD conditions require explicit treatment of device heterogeneity, echoing ubiquitous computing visions in which computation is embedded in everyday settings and must function amid mobility and heterogeneity [22,23]. Compatibility, permission variability, sensor quality variance, and OS fragmentation create a risk profile that is routinely addressed in enterprise mobility management and mobile security guidance. The North American National Institute of Standards and Technology (NIST) frameworks' guidance on managing mobile device security emphasizes configuration management, continuous monitoring, and operational controls, reflecting the broader point that mobile services require operational governance to remain dependable over time [3]. Although such guidance is security-oriented, it maps naturally onto broader service operability requirements relevant to city-scale AR in public settings, including update management, device readiness checks, and fallback pathways when devices fail.

In addition, public-space enactment introduces safety and supervision as constraints that influence system design. For example, a "complete task" functional requirement is incomplete without operability requirements specifying pacing buffers, safe stopping points, and interruption recovery, especially where attention and movement co-occur. In this sense, requirements engineering for city-scale AR must be treated as socio-technical and operations-aware rather than purely software-functional.

### *2.3. Traceability: Evidence-to-Requirement-to-Artifact*

Traceability is central for auditability and replication. The classical formulation of the requirements traceability problem emphasizes that requirements must remain linked to their origins and to downstream artefacts to support change management and accountability [12]. In a city-scale AR learning service, this need is amplified because determinants arise from multiple evidence streams: teacher-facing operational records, field observations, and log-derived feasibility descriptors. In the absence of trace links, the knowledge pertaining to implementation remains implicit and poses challenges for effective transfer.

The concept of traceability can typically be divided into stages of pre-requirements and post-requirements. Considering this, pre-requirements traceability establishes connections between requirements and both stakeholder and contextual sources. Regarding design and implementation features, and testing artifacts are associated with post-requirements traceability. A foundational overview of traceability fundamentals frames traceability as the ability to describe and follow the life of a requirement in both directions, supporting verification, compliance, and evolution [24]. For the present paper's scope, evidence-to-requirement traceability is particularly important because determinants are treated as requirements signals: each requirement must be justified by an evidence anchor that is independently inspectable.

Contemporary research suggests that traceability remains challenging in practice due to perceived workload, tool friction, and unclear immediate value, even though benefits for compliance and maintenance are recognized [15]. Additionally, the traceability literature indicates that pre-requirements traceability is comparatively underdeveloped relative to post-requirements traceability, despite its importance for accountability and audit [13]. Overall, this supports the study's conceptual guidance to operationalize traceability in a compact, identification-driven matrix that links Determinants, Requirements, Artefacts, and Evidence anchors, aiming to reduce ambiguity and improve inspectability.

### *2.4. Reproducibility and Transfer Kits in Applied Systems*

Reproducibility in applied computing is increasingly approached as an artefact and packaging problem, not merely a reporting problem. In systems research, replication beyond the originating team depends on access to executable artefacts and their supporting materials, including scripts, data schemas, configuration specifications, and procedural documentation. Association for Computing Machinery (ACM)'s Artifact Review and Badging Policy [25,26] formalizes these expectations by defining what qualifies as an artefact and by distinguishing levels of availability and evaluation. It also makes explicit that the credibility of results increases when artefacts are well documented, exercisable, and accessible [27].

Within the present study's context, a transfer kit can be defined as an operations-ready artefact package that supports enactment and maintenance under real-world constraints. For city-scale AR services, packaging must extend beyond code-level assets to include deployment artefacts such as marker production specifications, placement and maintenance procedures, onboarding scripts, incident-response routines, and device readiness checklists. This framing aligns with the broader reproducibility principle that what is reproduced is system behavior under specified conditions, which requires both technical artefacts and the operational procedures that stabilize those conditions.

Transfer kits also enable auditability. When requirements are linked to artefacts, and artefacts are versioned and packaged, inspection becomes feasible: a requirement can be traced to a specific operational checklist or technical module, and evidence anchors can be evaluated against the referenced artefact. Consequently, reproducibility-oriented packaging is treated here as a design requirement of the contribution, not an optional supplement [28,29].

### *2.5. Domain Constraints (Heritage, Curriculum, Governance) as Constraints*

City-scale AR learning services deployed in cultural heritage contexts operate under domain constraints that shape requirements without needing to dominate theoretical framing [30,31]. Heritage governance introduces stewardship, public awareness, and sustainability expectations that influence how points of interest are selected, how interpretive materials are licensed and maintained, and how public-space engagement is managed. At the European-context policy level, cultural heritage has been framed as a strategic resource for sustainable development, including education and public participation dimensions that affect municipal deployment conditions [32].

Curriculum framing operates as an adoption constraint: school uptake is more likely when activities are demonstrably alignable with learning objectives and legitimate within school routines. The European Sustainability Framework, also known as the GreenComp framework is a relevant reference point in this paper only insofar as it shapes how sustainability-related learning objectives are articulated and constrained in educational settings, without serving as the main axis of contribution [33]. Broader digital education policy and competence frameworks can be treated as contextual implementation constraints that shape teacher capacity, citizen digital competence, and the feasibility of mobile learning at scale, without becoming claims about learning outcomes in this study [34–38].

Governance and privacy constraints are similarly treated as boundary conditions. In European schools deployments, data minimization and lawful processing under General Data Protection Regulation (GDPR) constrain what telemetry can be collected and how data flows must be specified [39]. For this reason, privacy-aware implementation is positioned as a requirements driver, reinforced by privacy engineering guidance that promotes risk-based privacy management across the life cycle [17]. These constraints motivate group-level logging, explicit data retention policies, and purposed limitation as part of the service specification. The next section operationalizes these constraints as an explicit service boundary, evidence streams, and analysis procedures used for determinant quantification and requirements derivation.

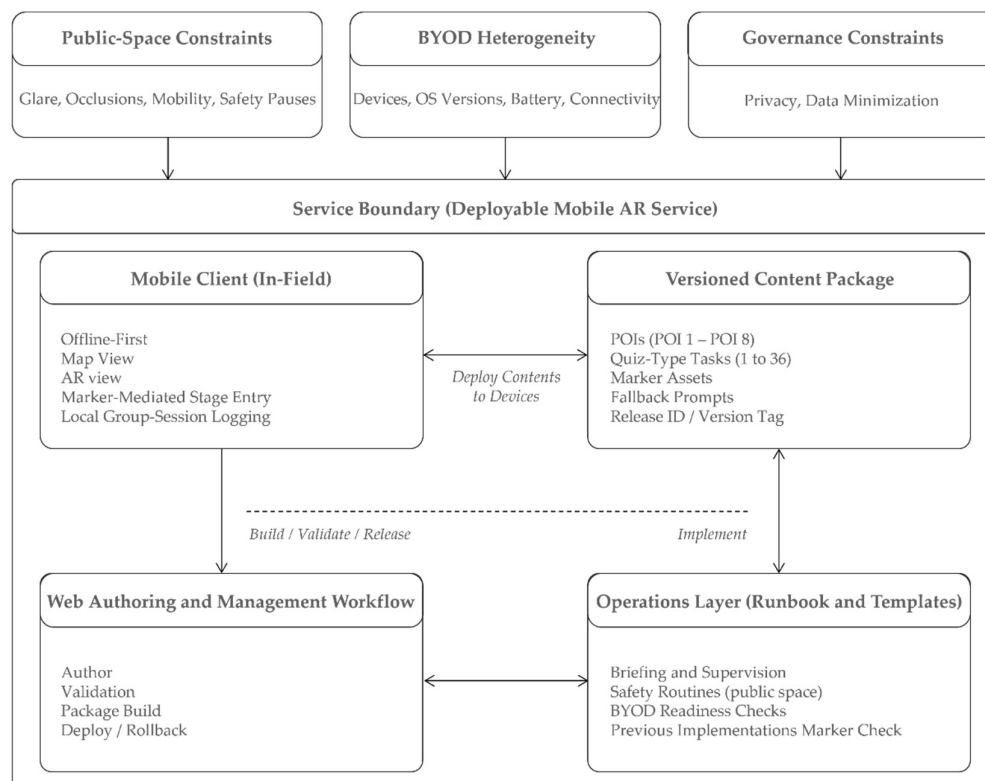
## **3. Materials and Methods**

### *3.1. System Overview and Boundary of the Service*

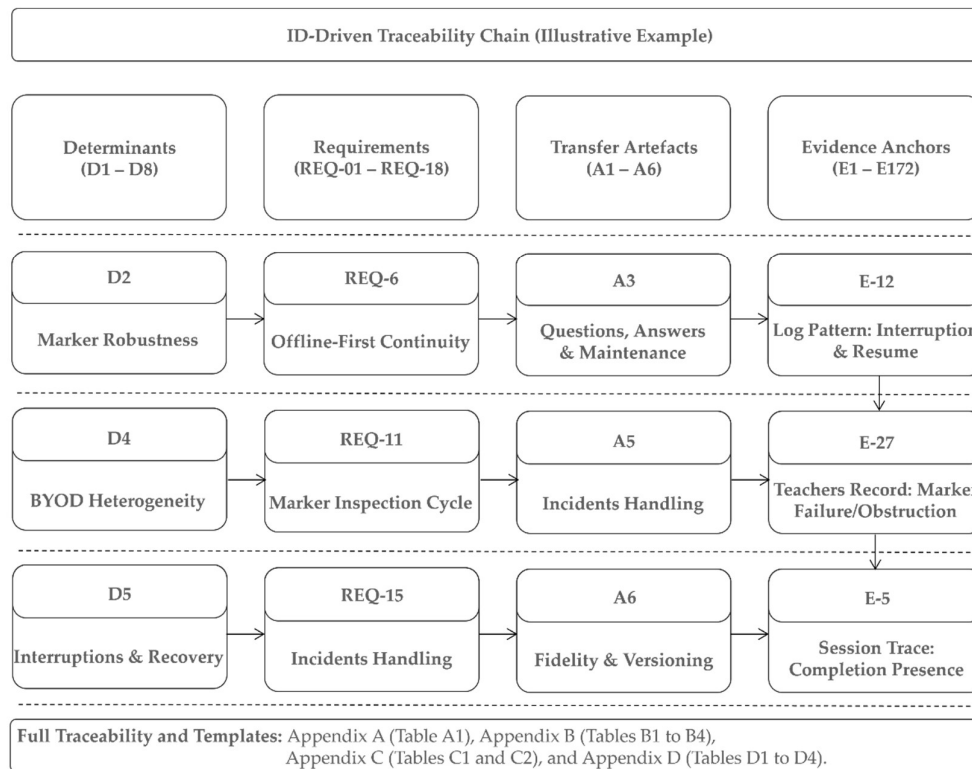
In this study, the *Art Nouveau Path* is analyzed as a deployable mobile AR service embedded in a local-scale ecosystem. The service boundary includes: (i) the mobile client used in the field, (ii) the authored task and content package associated with POIs and 36 tasks, (iii) a web-based authoring, the EduCITY's project website (available at: <https://educity.web.ua.pt/index.php>, accessed on 23 February 2026) and management workflow supporting content deployment, and (iv) an operations layer that stabilizes enactment under public-space constraints (briefing, supervision, safety routines, device readiness checks, and maintenance procedures) This service framing emphasizes enactment feasibility rather than learning outcomes.

Regarding replication, it requires rigorous management of versions features and operational artefacts. Content is authored and packaged into versioned releases that bundle POI definitions, marker assets, tasks, and fallback prompts, then deployed to field devices through the management workflow. During enactment, the client records group-session traces locally and these are harvested post-session for secure storage and for requirements traceability. Identifier namespaces are used consistently to link determinants, requirements, transfer artefacts, and evidence: Determinants (D1 to D8) denote coded implementation signals, Requirements (REQ-01 to REQ-18) denote derived specification items, transfer Artefacts (A1 to A6) denote reusable operational components, and Evidence anchors (E\_ID) denote audit-ready pointers to a source record identifier or log pattern supporting inspection. In this paper, instantiated anchors are enumerated as T1-VAL teacher records (E-T1VAL-R001 to E-T1VAL-R030), T2-OBS teacher records (E-T2OBS-R001 to E-T2OBS-R024), and Logs session records (E-LOG-R001 to E-LOG-R118), totaling  $n = 172$  (Appendix D, Tables A8–A10).

The service boundary and operations context are summarized in Figure 1, and the traceability chain is summarized in Figure 2, that presents an illustrative example of the ID-driven traceability mechanism, showing representative trace links and evidence anchor types rather than enumerating the complete set of instantiated anchors.



**Figure 1.** Service boundary of the *Art Nouveau Path* as a deployable city-scale, in-the-wild mobile AR service, including the field mobile client, versioned content package, web-based authoring and management workflow, and an operations layer stabilizing enactment under public-space constraints, BYOD heterogeneity, and governance constraints.



**Figure 2.** ID-driven traceability chain linking Determinants (D1 to D8) to derived Requirements (REQ-01 to REQ-18), transfer Artefact packs (A1 to A6), and instantiated Evidence anchors (E\_ID, n = 172). The example anchors labels shown in the figure are demonstrative and do not enumerate the full instantiated anchor set, which is listed in Appendix D (Tables A8–A10).

The full determinant-to-transfer kit traceability matrix, including determinant evidence totals and the associated transfer artefacts, is provided in Appendix A (Table A1) to enable audit and replication without reliance on external supplementary files.

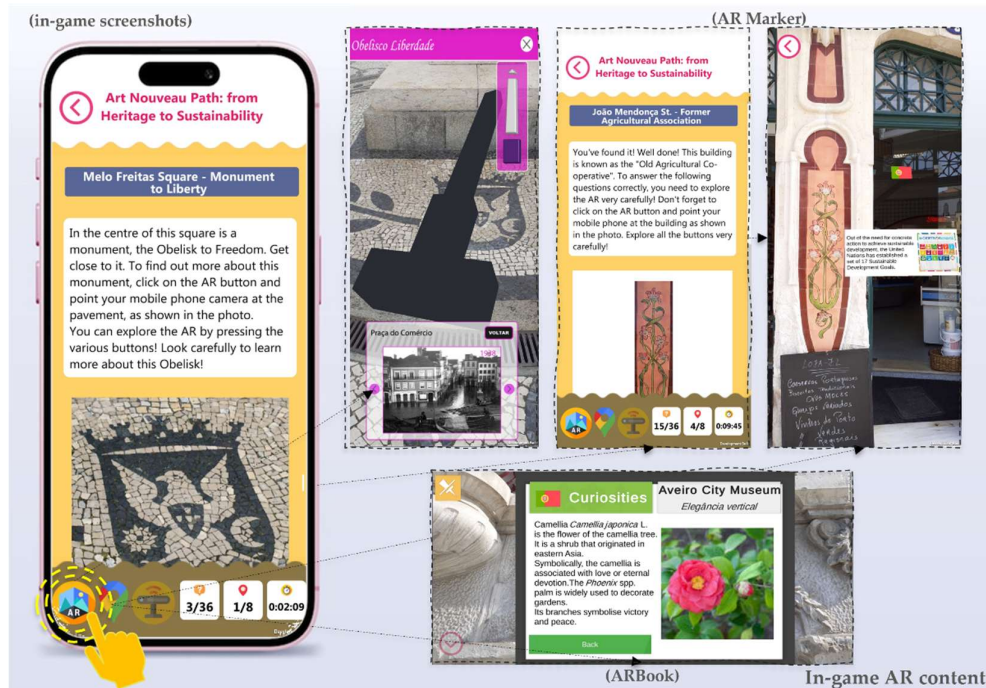
To support transfer beyond the originating team, the operational templates referenced in the requirements catalogue are included in Appendix B. These templates cover briefing and supervision checklists, marker inspection and maintenance logs, incident report forms, BYOD readiness checks, and fallback protocols, and are intended to make operational readiness inspectable and repeatable across deployments. In addition, to support inspection of the instrumentation strategy under governance constraints, the logging schema and a redacted example record are included in Appendix C, providing a sufficient specification for replication of the determinant-to-requirements derivation workflow [25,40,41].

To preserve empirical distinctiveness from competence-impact reporting within the broader research program, the analytical scope is intentionally restricted. Learning outcomes, psychometric modelling, and correctness-based performance inference are treated as out of scope of this study. Log and questionnaire artifacts are used only to describe feasibility envelopes and integrity constraints that bound deployment conditions and operations requirements [5,42].

### 3.2. Architecture and Instrumentation (Client, Backend, Content Pipeline, Logging)

The EduCITY's mobile application (available at: <https://educity.web.ua.pt/app.php>, accessed on 18 February 2026) (v1.3) is an offline-first Android and iOS client that combines location awareness (Global Positioning System (GPS) and compass) with image-based AR triggering via the *Vuforia SDK*. The primary interaction loop alternates between a 2D map view (path navigation, POI discovery) and

an AR camera view (marker scanning and AR activation). Two marker-mediated access modalities were used for stage entry in the path and in the profiling logs: ARBook and AR marker (Figure 3). Both labels denote marker-triggered access mechanisms used to open a stage (not the presence of authored AR overlays).



**Figure 3.** Marker-mediated stage entry modalities used in the *Art Nouveau Path* client and represented in the logs (ARBook and AR marker). Both labels denote marker-triggered access to a stage, not the presence of authored AR overlays.

In city-scale conditions, offline-first operation is treated as an operability requirement rather than a convenience: core task flow remains usable under intermittent connectivity, while deferred upload supports post-session data transfer and audit.

The content pipeline supports rapid updates without requiring continuous connectivity in the field. Authoring produces a structured package (POI metadata, task definitions, media assets, and marker descriptors) that is validated and deployed through the management workflow. Operationally, the package is treated as a versioned artefact so that deployments can be reproduced, audited, and rolled back when failures are detected, consistent with requirements lifecycle management and quality assurance practice [5,10].

Operational procedures are treated as first-class system artefacts because city-scale enactment involves predictable failures and recoveries. The operational stack includes briefing and supervision routines, marker inspection and replacement cycles, device readiness checks under BYOD heterogeneity, and incident handling for interruptions (app switching, tracking failure, safety pauses, and connectivity loss), aligning with service reliability practice for operating production systems [9]. The concrete runbook structure and template set used to operationalize these procedures are included in Appendix B to make transfer conditions inspectable and to reduce adoption variance across non-originating teams.

Gameplay traces are recorded at the group-session level because enactment is collaborative, typically one device per student group. Logs are captured locally on-device during the path and retrieved after each session for upload to a secure university server, with handling procedures consistent with information security management requirements and control guidance [43,44]. Records are group-level and designed to contain no direct personal identifiers, aligning with data

minimization constraints for school deployments in public space [17,39]. Although the logging schema contains fields that could support performance inference (for example, selected option and correctness), this paper restricts usage to feasibility descriptors: response presence, POI-level completion traces, and duration envelopes. The logging schema and a redacted example record are included in Appendix C to support auditability under governance constraints.

Instrumentation decisions were guided by two principles: (i) service auditability for requirements traceability, and (ii) data minimization proportional to the operational aim of transfer support. This aligns with reproducible computational research recommendations that emphasize recording how outputs are produced, retaining versioned artifacts, and preserving machine-readable intermediate outputs where feasible [40,41]. Where artifacts cannot be openly released due to governance constraints, Findable, Accessible, Interoperable, Reusable (FAIR)-aligned metadata and structured inventories remain applicable as a minimum for reusability and inspection [45,46].

### 3.3. Task Model and POI Profiling Procedure (36 Tasks, 8 POIs)

The *Art Nouveau Path* is implemented as eight geolocated POIs in Aveiro, Portugal, with 36 quiz-type tasks distributed across the POIs. Each POI functions as a compact challenge block: teams access a set of multiple-choice questions supported by authored multimodal resources (AR overlays, historical images, short video, audio, and text) before moving to the next location. A POI profiling procedure was applied to characterize deployment-critical dependency structure, specifically the extent to which task flow depends on marker-triggered interaction versus low-tech solvability. For each POI and for the path overall, tasks were classified by: (i) presence of authored AR overlay, (ii) whether access to the question stage required marker recognition, and (iii) whether the task solution demand was low-tech (observation or prior knowledge), even when the interaction layer required marker access. This profiling supports later contingency planning by identifying where marker dependence concentrates and where low-tech solution demand may provide resilience under outdoor variability [42,47].

### 3.4. Evidence Streams and Study Design

Evidence was organized into bounded streams with explicit analytical units and restricted roles to prevent analytical overlap with learning-impact reporting. Teacher-facing evidence (T1-VAL and T2-OBS) provides the primary basis for determinant identification and quantification, while logs provide feasibility descriptors only (completion traces and duration envelopes). The immediate post-path students' questionnaire (S2-POST) provides binary acceptability indicators and administration integrity descriptors; it is not used for learning-outcome inference. Specialist Teachers' narratives (T1-R; N = 3) are used only to contextualize transfer framing and to sanity-check requirement wording; they are not included in determinant coding or in the determinant-requirement-artefact-evidence traceability matrix. Table 1 summarizes these evidence streams, their analytical units, and their restricted roles in this study.

**Table 1.** Evidence streams, analytical units, and restricted analytical role in this manuscript.

Data Stream	Instrument/Source	N (records)	Analytical Unit	In-Study's Main Use
<i>Teacher's validation workshop questionnaire</i>	T1-VAL (open and closed fields)	30	Teachers' records; meaning units (open fields)	Adoption constraints; transferability criteria; determinant quantification

<i>Specialist curriculum review</i>	T1-R (expert narratives and heuristics)	3	Specialists' records; meaning units	Contextual triangulation only; not included in determinant coding or traceability matrix Public-space
<i>In situ teacher observation</i>	T2-OBS (structured observation and open fields)	24	Teachers' records; meaning units (open fields)	orchestration constraints; determinant quantification Marker-dependence profiling; contingency-relevant structure Feasibility envelopes: completion traces and duration descriptors only
<i>POI and task profiling</i>	Design documentation and content inventory	8 POIs; 36 tasks	POI-level dependency descriptors	Post-path acceptability constraints and administration integrity only, no outcome inference
<i>Group-session logs</i>	EduCITY app logs (group sessions)	118 sessions	Session-level event traces	
<i>Post-path students' questionnaire</i>	S2-POST (binary feasibility items and GCQuest block)	439	Binary feasibility indicators; questionnaire integrity descriptors	

Table 1 makes the analytical boundary explicit: teacher-facing records provide determinant signals, whereas logs and S2-POST are restricted to feasibility, acceptability constraints, and administration integrity descriptors, with no learning-outcome inference in this work.

Field enactment occurred through standardized routines: safety and interface briefing, path execution by students' groups, and post-path questionnaire administration. The field cohort comprised a convenience sample of 439 students (ages 13–18) distributed across 19 classes, with 118 valid collaborative group sessions present in the logs. A learners-per-session proxy was computed as  $439/118 = 3.72$ , treated strictly as an enactment descriptor rather than a precise group-size estimate.

### 3.5. Determinant Taxonomy D1–D8 (Operational Definitions)

An eight-code implementation determinant taxonomy (D1–D8) was used to translate heterogeneous teacher-facing evidence into actionable constraints and enablers for deployment and transfer. The taxonomy is mixed in origin: an initial codebook was informed by prior literature as sensitizing constructs and then refined through iterative coding and memoing to improve empirical fit, consistent with directed content analysis and hybrid inductive deductive thematic procedures [48–51]. The number of determinants was not fixed a priori. Eight categories represent a parsimony versus granularity trade-off: the set is sufficiently fine-grained to preserve distinct, implementation-actionable constraints, yet compact enough to sustain stable single-label assignment and reliable

multi-coder quantification for downstream requirements derivation. Codes are intentionally implementation-facing and capture operationally relevant issues rather than pedagogical quality judgments. Single-label assignment was enforced to support quantification and traceability. Table 2 defines the determinant codes (D1 to D8) and the operational cues used for single-label assignment.

**Table 2.** Implementation determinant taxonomy (D1–D8) with operational coding cues.

Code	Determinant (Primary Focus)	Operational Coding Cues (Inclusion Criteria)
D1	Curriculum alignment and framing	Curricular fit, disciplinary linkage, lesson framing, learning aims, legitimacy for school practice, integration in class
D2	Marker robustness and recovery	Marker recognition failures, AR trigger reliability, scanning issues, glare, positioning, recovery steps, alternative triggers
D3	Usability, legibility, and onboarding	Interface clarity, instructions, onboarding, task legibility, map/AR switching friction, confusion points, first-use support
D4	Post-activity consolidation and follow-up	Debrief needs, reflection prompts, consolidation packs, classroom follow-up, assessment logistics after the path
D5	Differentiation and accessibility	Accessibility requirements, inclusion, varied difficulty, support for diverse learners, readability and usability accommodations
D6	Safety and supervision in public space	Risk cues, supervision needs, safe stopping points, mobility constraints, attention switching, group control under movement
D7	Collaboration and accountability routines	Group roles, collaboration issues, coordination breakdowns, accountability, one-device-per-group dynamics and fairness
D8	BYOD heterogeneity and low-tech fallback	Device variability, compatibility issues, battery/network constraints, fallback routines, alternative access when devices fail

Table 2 operationalizes the determinant taxonomy as an inspectable coding instrument, defining inclusion cues that support consistent single-label assignment. The taxonomy supports requirements engineering by converting determinant signals into candidate requirements aligned with NFR categories such as reliability, usability, security, maintainability, and portability [42], and by making operational constraints explicit as part of the service specification [5].

### 3.6. Qualitative Coding Protocol and Reliability

Teacher-facing open-text evidence from T1-VAL (N = 30) and T2-OBS (N = 24) formed the qualitative corpus for determinant quantification. Analysis proceeded via meaning-unit segmentation, with each meaning unit assigned exactly one primary determinant code (D1–D8). When a meaning unit plausibly matched multiple determinants, precedence rules privileged implementation-critical constraints (for example, safety and supervision cues prioritized over usability when public-space risk was implicated; BYOD constraints prioritized over usability when

device capability was the limiting factor). Across 54 teachers' records, a total of 131 meaning units were extracted and considered. Meaning-unit counts are descriptive signals and are not treated as independent observations; teacher-record coverage is reported alongside meaning-unit totals to reduce overrepresentation from more verbose respondents. This design reduces double counting and improves downstream traceability.

Intercoder agreement was evaluated based on a stratified calibration subset designed to represent both corpora and to ensure coverage of lower-frequency determinants ( $n = 40$ ). The reliability for nominal, multi-category coding was then quantified using Krippendorff's alpha ( $\alpha = 0.83$ ; 95% bootstrap CI [0.72, 0.92]) alongside an exact-match criterion defined as unanimous agreement across all three coders (77.50%). The multi-coder content analysis and nominal classification tasks reporting follows established guidance [52,53]. Sampling parameters, coder outputs, and contingency summaries were preserved as an internal audit record for traceability. Also, the overall coding process followed an iterative, memo-supported workflow typical of applied qualitative analysis [50,51]. Data cleaning rules and denominators were fixed prior to analysis and retained as an internal audit record.

### 3.7. Quantitative Feasibility and Acceptability Descriptors from Logs and S2-POST

Log-derived descriptors were computed to bound the operational envelope of city-scale enactment. Three indicators were prioritized: (i) Response presence: whether a task recorded a response event (irrespective of correctness); (ii) Full path completion: response presence recorded for all 36 tasks within each session ( $N = 118$ ), operationalized as a total per-session response count equal to 36 (correct + incorrect); and (iii) Duration envelopes: session-level duration descriptors taken from the log export (minutes), computed as elapsed time within a session from interaction events. All log-derived outputs are treated descriptively; no inferential claims were drawn.

An immediate post-path student questionnaire (S2-POST) [54] was administered after path completion. It included (i) binary acceptability and feasibility items, (ii) optional open-text items, and (iii) the 25-item of the GreenComp-Based Questionnaire (GCQuest) block (Q1–Q25). In this manuscript, S2-POST is used only to (a) bound post-path acceptability constraints via the binary items, and (b) verify administration integrity via item-level completion and missingness in the GCQuest block. Missingness is summarized as complete-case records for the binary acceptability and feasibility items, complete-case records for Q1–Q25, and total missing cells across Q1–Q25.

### 3.8. From Determinants to Implementation Requirements

Requirements were derived by treating determinants as requirements signals. Determinant-coded meaning units from the teacher-facing qualitative corpus were reviewed to extract implementation-relevant constraints and translated into candidate requirements using a standardized template aligned with ISO's requirements guidance [5]. Semantically equivalent constraints were unified to minimize redundancy while retaining trace links to the contributing evidence pools and anchor inventories (Appendix D). Each requirement item (REQ-01 to REQ-18) includes: (i) Requirement statement: a clear, testable "shall" statement; (ii) Type: functional, quality attribute (non-functional), or operational requirement; (iii) Determinant source: linkage to the corresponding Dk code; (iv) Rationale: a brief justification grounded in the evidence anchor(s); (v) Acceptance criteria: observable verification conditions for field deployment; and (vi) Priority and criticality: assigned using a risk-aware heuristic that privileges safety-critical and enactment-blocking determinants, consistent with ISO 31000 risk management principles [47]. This procedure is consistent with established requirements engineering practice emphasizing elicitation from stakeholder evidence and the explicit handling of non-functional requirements as first-class requirements [42].

### 3.9. Traceability Model and Matrix Construction (Determinant-Requirement-Artifact-Evidence)

A traceability model was operationalized to support auditability and replication. Trace links were constructed across four levels: (i) Determinant (D1 to D8): coded implementation signal; (ii) Requirement (REQ-01 to REQ-18): derived specification item; (iii) Artefact (A1 to A6): transfer kit component or operational asset (for example, checklist, marker maintenance procedure, onboarding script, fallback protocol); and (iv) Evidence anchor (E\_ID): audit-ready pointer to a source record identifier or log pattern supporting inspection (Appendix D).

Traceability follows the bidirectional rationale of requirements traceability, enabling inspection from evidence to requirement and from requirement to implementation artefacts [12,55]. The matrix is designed to be compact in the manuscript body (high-level view) and extensible as a full supplementary artifact for operational transfer and audit. Traceability also supports reproducibility goals by making the rationale for each transfer artefact inspectable [41,56].

### 3.10. Ethics, Privacy-Aware Implementation, and Data Minimization

The deployment involved minors in school-organized activities and therefore requires explicit privacy-aware implementation. The logging strategy follows data minimization: group-level session identifiers are used; no direct personal identifiers are stored in the log dataset used for this manuscript; and telemetry is restricted to feasibility descriptors needed for operability and traceability. Considering the above, this study's data governance is shaped by the GDPR, namely regarding limiting purpose and minimizing data use [39]. Privacy risk management follows a risk-based approach consistent with the NIST Privacy Framework and privacy information management guidance, emphasizing proportional controls and documentation for operational accountability [17,57]. BYOD use-safety and operational readiness constraints are treated as part of the service's operational specification, consistent with enterprise guidance on managing mobile devices [3]. This boundary framing enables the results to be interpreted as service feasibility and specification signals rather than as learning-effectiveness claims.

## 4. Results

Results are structured to answer RQ1 to RQ3. First, profiling outputs establish the task and POI dependence structure that constrains enactment conditions. Second, determinant signals summarize implementation drivers and constraints (RQ1). Third, the derived requirements catalogue, traceability matrices, and minimal operations stack translate those signals into an auditable deployment specification (RQ2 and RQ3).

### 4.1. Task and POI Profiling Results

Task-architecture profiling was performed across the implemented eight-POI path (36 tasks) to characterize where city-scale enactment is structurally dependent on marker-triggered access and where resilience is supported through lower-dependency layers (for example, observation or knowledge-driven prompts). Table 3 summarizes POI-level dependency structure, separating (i) authored AR overlays, (ii) marker-triggered access concentrated at the question layer, and (iii) low-tech solution demand. The indicators present interdependencies and do not segment the 36 tasks.

**Table 3.** POI-level task dependency profile<sup>1</sup> across the eight-point path (N = 36 tasks): AR overlay tasks, marker-triggered question access, and low-tech solution demand (Observation (OBS) and Knowledge (KNOW)). Indicators are not mutually exclusive.

POI	Location	Total tasks (n)	AR overlay tasks (n)	AR overlay (%)	AR	Marker-triggered question tasks (n)	Marker-triggered question (%)	Low-tech	Low-tech (OBS and KNOW) (%)
					Marker-triggered question tasks (n)			tasks (OBS and KNOW) (n)	

	Joaquim de Melo							
1	Freitas Square: Obelisk to Liberty'	5	1	20.00	1	20.00	0	0.00
	Joaquim de Melo							
2	Freitas Square: 'Ala Pharmacy (old)'	4	2	50.00	1	25.00	2	50.00
3	João Mendonça Street	5	0	0.00	4	80.00	5	100.00
	João Mendonça							
4	Street: 'Old Agricultural Cooperative'	5	1	20.00	4	80.00	4	80.00
	João Mendonça							
5	Street: 'Aveiro City Museum'	6	3	50.00	6	100.00	4	66.67
6	'Art Nouveau Museum'	6	3	50.00	5	83.33	1	16.67
7	'José Estêvão Market' (Fish Market)	3	1	33.33	3	100.00	3	100.00
8	'Ferro Guesthouse'	2	0	0.00	2	100.00	2	100.00
	Totals	36	11	30.56	26	72.22	21	58.33

<sup>1</sup> Note: The three profiling indicators are not mutually exclusive; a task may require marker access while still being low-tech solvable, so counts across indicators should not be interpreted as partitioning the 36 tasks.

Table 3 reveals two structural patterns that matter for operability and contingency planning, namely selective dependency, since authored AR overlays represent 30.56% of tasks, while marker-triggered question access is required in 72.22% of tasks. City-scale robustness is therefore gated less by “overlay intensity” and more by stabilizing marker-triggered access and recovery at the question layer. Regarding POI modularity with local closure, tasks are packaged into POI-bounded micro-sequences, supporting completion on-site before mobility to the next location. This packaging reduces the propagation of disruptions during interruptions among POIs. Also, stage-level profiling further clarifies where marker-triggered access concentrates across the four-stage sequence (Intro cue, Question, Correct feedback, Incorrect feedback). “ARBook” and “AR marker” denote marker-triggered access mechanisms used to open a stage, not the presence of authored AR overlays.

Table 4 reports both counts and percentages per stage.

**Table 4.** Primary modality by task stage (dominant modality per stage; N = 36 tasks per stage).

Stage	AR	AR	Text		Image		ARBook		Audio		Video	
	Marker	Marker	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)
	(n)	(%)										
<i>Intro cue</i>	1	2.78	5	13.89	7	19.44	19	52.78	1	2.78	3	8.33

<i>Question</i>	1	2.78	9	25.00	1	2.78	25	69.44	0	0.00	0	0.00
<i>Correct feedback</i>	1	2.78	23	63.89	10	27.78	0	0.00	1	2.78	1	2.78
<i>Incorrect feedback</i>	1	2.78	21	58.33	12	33.33	0	0.00	1	2.78	1	2.78

As stated in Table 4, marker-triggered access concentrates in the Question stage (ARBook and AR marker = 26/36 stages, 72.22%), while both feedback stages are predominantly text-mediated. Operationally, this indicates that the most technically sensitive moment is opening the question layer, whereas closure is intentionally low-dependency and thus more tolerant to mobility-related interruptions. At the same time, the concentration of text-mediated feedback elevates first-use legibility and readability constraints for learners and teachers under outdoor conditions.

#### 4.2. Log-Derived Feasibility Envelope (N = 118 Group Sessions) and Post-Path Feasibility Checks

Collaborative group-session logs were used exclusively as feasibility descriptors for city-scale enactment (completion traces, duration envelopes), not as learning-effectiveness indicators. Across valid logs, the total per-session response count equals 36 (correct + incorrect) for all N = 118 sessions, indicating full path completion under the completion criterion defined in Section 3.7. Table 5 summarizes the feasibility envelope from valid group-session logs (N = 118), including completion and time budgeting descriptors.

**Table 5.** Feasibility envelope from group-session logs (session-level descriptors; N sessions = 118).

Descriptor	Value
<i>Valid logged group sessions</i>	118
<i>Full path completion (sessions)</i>	118 (100.00%)
<i>Duration range (minutes)</i>	26.00 to 55.00
<i>Duration mean (minutes)</i>	42.38
<i>Duration median (minutes)</i>	42.00
<i>Duration IQR (minutes)</i>	38.00 to 45.80
<i>Learners per logged group session (proxy)</i>	3.72 (439 students / 118 sessions)

As presented in Table 5, across valid logs, total duration ranged from 26.00 to 55.00 minutes (Median (MDN) = 42.00; M (Mean) = 42.38; IQR (Interquartile range) = 38.00 to 45.80).

Post-path students' questionnaire indicators were used to bound acceptability and adoption-relevant assumptions under city-path conditions. These indicators are self-reported and are treated as feasibility constraints rather than effectiveness evidence.

The key acceptability and feasibility indicators (binary) are reported in Table 6.

**Table 6.** Student acceptability and feasibility indicators (post-path student's questionnaire; N = 439).

Indicator	Yes (n)	Yes (%)	No (n)	No (%)
<i>Interest in learning about sustainability through Art Nouveau heritage</i>	432	98.41	7	1.59
<i>Interest in learning more about Aveiro's Art Nouveau heritage</i>	414	94.31	25	5.69
<i>Self-reported ability to name sustainability competences</i>	265	60.36	174	39.64
<i>Perception that the game addresses sustainability competences</i>	434	98.86	5	1.14
<i>Perceived importance of sustainability competences</i>	427	97.27	12	2.73
<i>Interest in learning more about sustainability competences</i>	369	84.05	70	15.95

A key transfer constraint emerges in Table 6: despite near-ceiling endorsement for relevance and perceived competence-addressing, only 60.36% self-report an ability to name sustainability competences. Adoption should therefore not presume stable conceptual articulation without teacher mediation and post-activity consolidation.

Feasibility of immediate post-path questionnaire (S2-POST) administration is supported by near-complete completion of the 25-item GCQuest block. Table 7 details completeness and item-level missingness.

**Table 7.** Post-path students' questionnaire (S2-POST, N = 439); GCQuest block completeness (Q1 to Q25; N = 438); N total = 439; N complete-case = 438.

Indicator	Value (N/n and %)
<i>Complete-case records (all binary acceptability and feasibility items)</i>	439/439 (100.00)
<i>Complete-case records (all Q1 to Q25 present)</i>	438/439 (99.77)
<i>Total missing item cells (Q1 to Q25)</i>	7/10,975 (0.06)

As reported in Table 7, item-level missingness is negligible and concentrated in a single respondent (Q11 to Q17, 7 missing cells), supporting the interpretation that standardized immediate post-path administration is operationally feasible at cohort scale.

#### 4.3. Teacher-Facing Implementation Signals and Determinant Concentrations (T1-VAL and T2-OBS; Teacher Records N = 54)

Teacher-facing evidence provides implementation signals for service readiness and transfer packaging. Validation signals (T1-VAL) indicate high perceived feasibility for recommendation and curricular integration (Table 8).

**Table 8.** Teachers' validation signals (T1-VAL; N = 30).

Indicator	Yes (n)	Yes (%)	No (n)
<i>Would recommend to other teachers</i>	28	93.33	2
<i>Consider it feasible to integrate in curricular practice</i>	27	90.00	3
<i>Consider the tasks understandable without prior AR training</i>	27	90.00	3
<i>Intend to use the resource in future activities</i>	28	93.33	2

These validation signals (Table 8) may indicate strong adoption intent and perceived feasibility, supporting their use as determinant evidence for requirements derivation in subsequent tables.

Curricular dispersion (Table 9) indicates interdisciplinary uptake potential rather than single-subject confinement.

**Table 9.** Teachers' curricular and subject areas (T1-VAL; N = 30).

Curricular Area/Subject	n	%
<i>Civic Education</i>	6	20.00
<i>Arts</i>	5	16.67
<i>Geography</i>	5	16.67
<i>Mathematics</i>	5	16.67
<i>Multidisciplinary</i>	4	13.33
<i>History</i>	3	10.00
<i>Science</i>	2	6.67

Table 9 indicates interdisciplinary uptake potential, with Civic Education, Arts, Geography, Mathematics, and multidisciplinary framing accounting for most endorsements, supporting broad curriculum positioning.

In situ observation (T2-OBS, N = 24) provides orchestration-relevant feasibility indicators under mobility constraints. Ratings indicate consistently high endorsement for repeat participation and integration, while instruction clarity exhibits comparatively higher dispersion (Table 10).

**Table 10.** Key T2-OBS feasibility-related Likert indicators (1 to 6 scale; N = 24).

Item	Mean (M)	Standard Deviation (SD)	Median (MDN)	Min.	Max.
<i>Instructions were clear</i>	4.67	0.96	5	3	6
<i>Would participate again</i>	5.75	0.44	6	5	6
<i>Feasible to integrate in school practice</i>	5.08	0.58	5	4	6
<i>Appropriate across multiple grade levels</i>	4.88	0.61	5	4	6
<i>Perceived innovativeness of the resource</i>	5.62	0.49	6	5	6

Table 10 shows near-ceiling feasibility endorsement under in situ conditions, while instruction clarity remains comparatively lower, motivating explicit onboarding and legibility supports.

Binary enactment indicators (Table 11) support the interpretation that collaboration and problem solving are frequently observed. Observed exploration beyond the planned path is lower, suggesting that transfer beyond the designed path is not automatic and likely requires explicit prompting and mediation.

**Table 11.** Observed enactment indicators (Yes/No) (T2-OBS; N = 24).

Observation indicator	Yes (n)	Yes (%)	No (n)
<i>Activity supports exploring other places or paths</i>	15	62.50	9
<i>Activity supported discussion about sustainability</i>	16	66.67	8
<i>Activity supported care for public space</i>	20	83.33	4
<i>Activity supported relation to classroom content</i>	17	70.83	7
<i>Activity supported problem solving</i>	20	83.33	4
<i>Activity supported group collaboration</i>	18	75.00	6

Table 11 indicates strong observed collaboration, problem solving, and care for public space, whereas exploration beyond the planned path is less consistent, implying that off-path transfer should be explicitly scaffolded.

Open-field improvement focuses emphasize pragmatic orchestration and robustness constraints (Table 12). BYOD preparation and fallback planning dominate as the most frequent improvement category. Counts reflect the number of teacher records mentioning each category, and a single record may contribute to multiple categories. By contrast, Table 13 reports single-label determinant coding at meaning-unit level, so multi-topic statements are allocated to one primary determinant according to the precedence rules described in Section 3.6.

**Table 12.** Open-field improvements derived from T2-OBS suggestions (T2-OBS; N = 24). Categories are not mutually exclusive.

Category	Example improvement focus	Count (n/N)
----------	---------------------------	-------------

<i>Technical robustness and device constraints</i>	BYOD preparation, connectivity planning, low-tech alternative	14/24
<i>Orchestration and Group Management</i>	Cooperative inter-group challenges, time and pacing guidance	5/24
<i>Instruction legibility and teacher-facing scripts</i>	Teacher's guide, scripts for assessment and follow-up	4/24
<i>Differentiation and Accessibility</i>	Adaptations by age, Scaffolding	3/24
<i>Content Enrichment</i>	More contextual information, and additional heritage facts	3/24

**Table 13.** Implementation determinants and quantitative descriptors (single-label coding; MU N = 131; teacher records N = 54).

<b>Implementation determinant</b>	<b>Total MU (N)</b>	<b>T1- VAL MU (n)</b>	<b>T2- OBS MU (n)</b>	<b>Teacher records mentioning (n/N)</b>	<b>Teachers mention (%)</b>	<b>MU (%)</b>	<b>Transfer kit component</b>
<i>D3: Usability, legibility, onboarding</i>	29	22	7	25/54	46.30	22.14	Teacher-facing quick start; in-app legibility supports; onboarding notes
<i>D2: Marker robustness and recovery</i>	25	14	11	22/54	40.74	19.08	Marker deployment guidance; recovery steps; alternative triggers
<i>D1: Curriculum alignment and framing</i>	24	19	5	21/54	38.89	18.32	Curriculum mapping matrix; facilitation and framing script
<i>D4: Post-activity consolidation</i>	14	9	5	13/54	24.07	10.69	Structured debrief template; classroom follow-up prompts
<i>D5: Differentiation and accessibility</i>	14	9	5	12/54	22.22	10.69	Adaptation variants by age; accessibility notes
<i>D8: BYOD heterogeneity and fallback</i>	10	4	6	9/54	16.67	7.63	Device preparation and compatibility checks; low-tech fallback options
<i>D6: Safety and supervision</i>	8	4	4	7/54	12.96	6.11	Safety briefing; supervision and public-space cues
<i>D7: Collaboration and accountability routines</i>	7	3	4	6/54	11.11	5.34	Role cards; device-sharing protocol; regrouping scripts

Table 12 shows that improvement requests concentrate on BYOD and robustness constraints, with secondary needs in orchestration scripts and differentiation, reinforcing the determinant emphasis that follows.

To consolidate teachers-facing evidence into an implementation taxonomy, open-field content across teacher records (T1-VAL and T2-OBS; total teacher records N = 54) was coded into meaning units (MU) under a determinant taxonomy (D1 to D8). Table 13 reports determinant concentrations, including MU totals, split by teacher corpus, record coverage, and the corresponding transfer kit component.

As presented in Table 13, three determinants dominate both MU frequency and record coverage: D3 (usability, legibility, onboarding), D2 (marker robustness and recovery), and D1 (curriculum alignment and framing). In combination, these determinants function as implementation gating factors for first adoption and repeatable enactment. Lower-frequency determinants (D6 to D8) are treated as enactment-critical because they are coupled to public-space risk and BYOD fragility, and therefore enter the requirements catalogue as operational constraints rather than optional enhancements.

A compact synthesis of these teacher-facing implementation signals is provided in Table 14 to support rapid inspection in transfer contexts.

**Table 14.** Teacher-facing feasibility and implementation signals (summary view).

Source	Indicator Type	Key Result (Descriptive)
T1-VAL (N = 30)	Recommendation and intent	High endorsement for recommending and future use
T1-VAL (N = 30)	Instruction clarity	Lower dispersion than technical concerns, but variability remains at first-use
T2-OBS (N = 24)	Enactment constraints	Recurrent needs in safety routines, pacing buffers, and group orchestration
T2-OBS (N = 24)	Improvement requests	Concentration in robustness, BYOD constraints, and teacher-facing scripts

Together, the summary signals in Table 14 show convergent feasibility endorsement alongside specific first-use friction points, which motivates the shift from descriptive evidence to explicit, testable deployment Requirements.

#### 4.4. Derived Requirements Set (REQ-01 to REQ-18), Grouped by Determinant

Determinant evidence was converted into auditable “shall” statements mapped to transfer artefacts. Statement structure follows requirements specification practice in ISO’s requirements [5]. Requirements are grouped by Determinant for auditability and prioritization, and expressed as a minimal operations-ready set.

Table 15 lists the minimal operations-ready requirements set (REQ-01 to REQ-18) with verification cues and transfer artefact mappings.

**Table 15.** Determinant-driven requirements catalogue (minimal operations-ready set).

REQ ID	Determinant	Type <sup>2</sup>	Requirement statement (shall)	Acceptance criteria (verification)	Transfer artefact(s)
REQ-01	D1	OP	Provide a curriculum-to-task mapping matrix covering all POIs and tasks.	Matrix includes all POIs and tasks with explicit curriculum descriptors and intended learner outputs.	A2

<i>REQ-02</i>	D1	OP	Provide a teacher-facing facilitation and framing script for enactment.	Script specifies learning aims, time budget, group roles, expected outputs, pacing guidance, and closure prompts (1 to 2 pages).	A2
<i>REQ-03</i>	D3	OP	Provide a teacher-facing quick-start guide for first-time use.	One-page start routine plus core navigation cues; includes a minimal troubleshooting checklist. Onboarding notes address	A1
<i>REQ-04</i>	D3	OP	Provide onboarding notes that reduce first-use confusion.	scanning posture, path flow, and the distinction between question access and AR overlays.	A1
<i>REQ-05</i>	D3	NFR	Provide in-app legibility supports suitable for outdoor conditions.	Field check confirms instruction clarity under mobility and glare; font sizing and contrast cues are explicitly addressed. Deployment guide specifies	A1
<i>REQ-06</i>	D2	NFR	Provide marker production and deployment guidance suitable for outdoor use.	print spec, size, mounting, inspection points, and glare mitigation steps; replacement criteria are defined. Recovery protocol includes	A4
<i>REQ-07</i>	D2	FR	Provide explicit recovery steps for recognition failure and interrupted progression.	rescan strategy, repositioning, restart, rejoin, and teacher override steps; recovery is executable on-site. At least one alternative	A4
<i>REQ-08</i>	D2	FR	Provide alternative triggers or progression cues to reduce brittle marker dependence.	access path is defined per POI block (for example, teacher override, skip mechanism, or offline prompt).	A4
<i>REQ-09</i>	D8	OP	Provide BYOD readiness and compatibility checks.	Pre-session checklist covers device readiness, camera permissions, storage, battery, and connectivity;	A4

				common failure states are enumerated.	
<i>REQ-10</i>	D8	OP	Provide low-tech fallback options to sustain continuity.	Fallback includes non-AR progression cues and an offline or no-phone alternative for restricted contexts; materials are printable.	A4
<i>REQ-11</i>	D6	OP	Provide a safety briefing and public-space supervision cues.	Safety script includes supervision rules, crossing routines, and stop criteria; responsibilities are assigned before launch.	A3
<i>REQ-12</i>	D6	OP	Provide regrouping scripts and pacing buffers.	Routines include headcounts and buffer time guidance (for example, 5 to 10 minutes for crossings and regrouping).	A3
<i>REQ-13</i>	D7	OP	Provide role cards supporting accountability in group use.	Role cards specify responsibilities (navigator, scanner, recorder, timekeeper) and rotation rules.	A3
<i>REQ-14</i>	D7	OP	Provide a device-sharing protocol for equitable participation.	Protocol defines rotation frequency and ensures each learner accesses core interaction moments; accountability checks are included.	A3
<i>REQ-15</i>	D4	OP	Provide a structured debrief template for immediate consolidation.	Template includes prompts for reflection, evidence use, and sustainability framing; outputs are defined (oral, worksheet, or digital).	A5
<i>REQ-16</i>	D4	OP	Provide classroom follow-up prompts for post-path use.	Follow-up prompts include extension tasks aligned with curriculum descriptors and sustainability competences.	A5
<i>REQ-17</i>	D5	OP	Provide adaptation variants by age and ability.	Variants include simplified and extended pathways,	A6

				timing adjustments, and scaffolding suggestions.	
				Notes address mobility, sensory constraints, and alternative participation roles; inclusive design cues are provided.	
<b>REQ-18</b>	D5	OP	Provide accessibility notes addressing inclusion constraints.		A6

<sup>2</sup> Note: Functional Requirement (FR); Non-Functional Requirement (NFR); Operational Requirement (OP).

Overall, as presented in Table 15, the catalogue emphasizes operability: most requirements specify executable procedures, checklists, and recovery pathways rather than new features, reflecting the service boundary and the dominance of outdoor robustness and first-use onboarding constraints in the evidence.

#### 4.5. Requirements-to-Artifact Traceability and Determinant-to-Transfer Mapping

Traceability is operationalized as an inspectable mapping from determinants to requirements and transfer artefacts, enabling replication teams to audit why a given component exists and which determinant evidence it covers. The full determinant-to-transfer kit traceability matrix (absolute counts and component lists) is provided in Appendix A (Table A1). For rapid inspection, Table 16 summarizes determinant coverage across the requirements catalogue (REQ-01 to REQ-18) and the transfer artefact packs (A1 to A6).

**Table 16.** Compact determinant-to-requirement-to-artefact traceability (minimal operations-ready view).

Determinant	Primary REQs	Transfer Artefact(s)
<i>D1</i>	REQ-01, REQ-02	A2
<i>D2</i>	REQ-06, REQ-07, REQ-08	A4
<i>D3</i>	REQ-03, REQ-04, REQ-05	A1
<i>D4</i>	REQ-15, REQ-16	A5
<i>D5</i>	REQ-17, REQ-18	A6
<i>D6</i>	REQ-11, REQ-12	A3
<i>D7</i>	REQ-13, REQ-14	A3
<i>D8</i>	REQ-09, REQ-10	A4

As presented in Table 16, the minimal operations-ready reduce ambiguity about what a replication team must prepare, enact, and maintain, while preserving a path back to the determinant evidence totals reported in Section 4.3 and Appendix A.

#### 4.6. Minimal Operations Stack and Transfer Kit Outputs

Operational templates and runbook structures are provided in Appendix B (Tables A2–A5). These artefacts translate requirements into executable routines covering onboarding, safety and supervision, BYOD readiness checks, marker inspection and maintenance, incident handling, and fallback activation.

Across the teacher evidence corpus and the MU profiling evidence, enactment-critical determinants repeatedly surfaced as coordination, pacing, and recovery problems in the field. Accordingly, the minimal operations stack is expressed as a small set of executable loops covering session briefing and start-up, in-field supervision and safe pausing, post-session closure and secure data transfer, and routine maintenance with marker inspection and replacement. Appendix B makes these loops inspectable through an orchestration checklist (Table B1), a transfer kit artefact inventory

(Table A3), a minimal RACI for operational role clarity (Table A4), and routine definitions paired with evidence outputs (Table A5).

These artefacts are designed to minimize replication workload while preserving coverage of enactment-critical determinants. In particular, they separate conditions stabilized primarily by procedure (pacing buffers, accountability routines, interruption recovery) from those stabilized primarily by technical artefacts (marker packs, fallback prompts, and versioned releases). Data handling is specified as privacy-aware and proportional: only group-level traces are processed, no direct personal identifiers are required, and post-session harvesting and secure storage are treated as operational requirements consistent with GDPR minimization and risk-based mobile security [3,17,39].

## 5. Discussion

### 5.1. Determinants as Non-Functional and Operability Drivers in City-Scale AR Services

As presented in Section 4, the implementation signals are concentrated in D3 (usability and onboarding), D2 (marker robustness and recovery), and D1 (curriculum framing), with additional enactment-critical constraints in D6 to D8 (safety, collaboration routines, BYOD and fallback). These distributions are consistent with a service framing for in-the-wild AR, where deployment success depends less on feature breadth and more on whether interaction can be reliably initiated, resumed, and completed across variable outdoor conditions and heterogeneous devices [1,4].

From a software quality perspective, D2 and D3 map directly onto reliability and usability drivers that are typically treated as non-functional requirements, but become operationally dominant when the system boundary is expanded to include public-space mobility, device sharing, and non-expert facilitation [5,42]. In particular, D2 aligns with recoverability and fault tolerance expectations in outdoor AR workflows, given known sensitivities of AR tracking and marker-based access to lighting, occlusion, and environmental degradation [2].

The presence of a strong D3 signal indicates that first-use legibility and onboarding are not cosmetic improvements but gating conditions for adoption. In-the-wild interaction work has long emphasized that uncontrolled contexts amplify breakdowns that are absent or underrepresented in lab or pilot settings [1]. For interruption-prone mobile activity, this extends to the design of explicit resumption support, which is conceptually consistent with the recovery and fallback requirements derived in Section 4 [20].

To make the determinant logic inspectable for a computing audience, Table 17 provides a compact mapping from the determinant taxonomy to software-quality dimensions and to the artefacts reported in Section 4.

**Table 17.** Determinants as quality drivers and operability constraints (interpretive mapping grounded in Sections 3 and 4 evidence).

Determinant	Dominant quality driver (illustrative)	Operability implication	Trace output in this study
<i>D1 Curriculum alignment and framing</i>	Appropriateness and relevance	Framing as boundary condition, not learning effect claim	REQ-01 to REQ-02; Tables 9 to 10
<i>D2 Marker robustness and recovery</i>	Reliability and recoverability	Recovery runbooks, alternative triggers, maintenance cycle	REQ-06 to REQ-08; Tables 14 to 15

<b>D3 Usability and onboarding</b>	Usability and learnability	Quick-start, field legibility checks, facilitation scripts	REQ-03 to REQ-05; Tables 8 to 10
<b>D4 Post-activity consolidation</b>	Continuity across contexts	Debrief templates and follow-up prompts	REQ-15 to REQ-16; Table 14
<b>D5 Differentiation and accessibility</b>	Accessibility and inclusiveness	Alternative enactment variants	REQ-17 to REQ-18; Table 14
<b>D6 Safety and supervision</b>	Quality in use (risk reduction)	Crossing routines, regrouping, pacing buffers	REQ-11 to REQ-12; Tables 8, 14 to 15
<b>D7 Collaboration and roles</b>	Operability in group enactment	Role cards, turn-taking, accountability protocol	REQ-13 to REQ-14; Tables 6 to 8
<b>D8 BYOD heterogeneity and fallback</b>	Portability and compatibility	Device triage, offline or no-phone fallback	REQ-09 to REQ-10; Tables 7, 14 to 15

The mapping in Table 17 also makes an important boundary condition visible: determinants are treated as deployment and operations drivers that map to quality attributes and artefacts, not as proxies for educational effects.

### 5.2. Positioning Against Related Work in in-the-Wild AR, Requirements Engineering, and Traceability

The results reported are not novel AR techniques. This study concerns a deployment-oriented specification that treats the path as an educational software service with explicit operational constraints. Therefore, this study aims to answer an implementation-specification gap in MARG research by contributing a determinant-driven, transferable requirements engineering and traceability framework that links in-the-wild evidence to testable requirements and operations-ready artefacts for auditable replication [1,5,12,13,15]. Considering this framing, breakdowns, awkwardness, and context-driven deviations are treated as first-class evidence for specification and transfer, consistent with in-the-wild HCI accounts of how uncontrolled contexts surface failure modes that remain underrepresented in lab or pilot settings [1,4]. This contribution is situated within a broader program of research centered on the *Art Nouveau Path* and the EduCITY Digital Teaching and Learning Ecosystem (DTLE) [58], and it complements earlier design [59] and validation reporting [60] and log-based analytics instrumentation [61] by focusing specifically on transfer, auditability, and replication-ready packaging.

Within requirements engineering, the approach operationalizes teacher-facing determinants as requirements signals. This is consistent with foundational RE arguments that requirements emerge from socio-technical constraints and stakeholder contexts, and that explicit traceability is needed to make rationale inspectable and maintainable across system evolution [5,56].

The traceability model (Determinant to Requirement to Artefact to Evidence) implements the pre-requirements perspective summarized in Section 2.3 by linking requirements to inspectable origins in stakeholder records and operational observations [13]. The design targets maintainability and auditability through a compact, identification-driven matrix and explicit role clarity [15,55].

At the broader traceability level, this positioning remains consistent with the classical “requirements traceability problem” definition and its bidirectional linking logic [12].

### 5.3. Generalization Logic and Boundary Conditions

Generalization is intentionally constrained to deployment feasibility and operability, not learning impact. Robust implementation is a prerequisite for interpreting learning outcomes in city-scale MARG deployments because instability in implementation fidelity confounds pedagogy with delivery failures. When adherence to the intended flow, exposure dosage, and quality of delivery

vary due to breakdowns, interruptions, or operational gaps, outcome estimates become ambiguous and may reflect a mismatch between the intended and the delivered intervention, a well-described threat to interpretability associated with multidimensional fidelity constructs [62–64]. This concern is amplified in complex, real-world interventions, where updated guidance emphasizes the need to examine implementation, mechanisms, and context through process evaluation before attributing effects to the intervention [63,65]. For this reason, the present study prioritizes specification, traceability, and transfer artefacts that stabilize enactment conditions and make them inspectable, while explicitly accommodating the fidelity–adaptation balance expected in field deployments [66,67].

The findings support analytic generalization: determinant concentrations and the derived requirements highlight recurrent adoption gates and failure modes likely to arise in city-scale outdoor AR services, including marker fragility, lighting and glare effects on tracking, interruption recovery, and device heterogeneity [2,4].

Several boundary conditions delimit transfer:

First, Urban morphology and POI affordances: Variations in line-of-sight, pedestrian dynamics, and regrouping point accessibility influence the design and implementation of safety protocols, pacing adjustments, and interruption management across urban environments.

Second, Marker governance and maintenance capacity: Where inspection and replacement cycles cannot be sustained, marker-based access becomes a structural risk. In such contexts, alternative triggers and low-tech continuity tasks become higher-priority mitigations.

Third, school policy and device rules: Constraints on smartphone availability, permissions, or connectivity can narrow the effective BYOD envelope, shifting effort from optimization toward fallback orchestration and operational contingency.

Fourth, curriculum and institutional framing: Curriculum alignment functions as an adoption constraint rather than as evidence of learning outcomes. Where policy frameworks differ, the mapping artefact must be regenerated, while the determinant and traceability logic can remain stable.

For curriculum and sustainability framing, GreenComp [33] is treated as an alignment reference and contextual constraint rather than as an outcomes claim, consistent with the study boundary stated in Section 3.

#### *5.4. Practical Implications for Adoption, Replication, and Responsible Operations*

The most actionable implication of Sections 4.3 to 4.6 is that transfer should be treated as an operations problem, not only a content packaging problem. The determinant-driven transfer kit converts teacher-facing adoption signals into inspectable artefacts (quick-start, safety script, BYOD checklist, marker maintenance routine) that reduce first-use ambiguity and provide recovery paths.

This packaging logic is consistent with reproducibility guidance that emphasizes versioning, explicit workflows, and access to the artefacts that generate reported results [40,41]. It also aligns with data stewardship norms for discoverability and reuse when artefacts are shared in stable repositories with clear identifiers [46] and with computing-community expectations for artifact availability and auditability [25,26].

From a privacy-aware implementation standpoint, treating the service as deployable implies a default stance of minimization and clear operational roles for data handling, which is consistent with GDPR obligations for purpose limitation and data minimization in EU contexts [39].

To make adoption outcomes interpretable without overclaim, feasibility evidence in Sections 4.2 and 4.3 can be read through an “implementation outcomes” lens (acceptability, adoption intention, feasibility), while maintaining the explicit separation from learning outcomes, consistent with implementation-science distinctions [7].

## **6. Conclusions, Limitations, and Future Paths**

### *6.1. Conclusions and Summary of Contributions (C1–C4)*

This paper reframed the *Art Nouveau Path* as a deployable city-scale, in-the-wild mobile AR learning service, where operational success depends on service-level reliability and governance rather than on isolated application features. Consistent with in-the-wild HCI perspectives, breakdowns and repairs were treated as expected conditions of use rather than exceptional events [1,4]. Evidence from task and POI profiling, collaborative group-session logs, and teacher-facing records was used to derive an auditable engineering specification focused on feasibility, operability, and transfer. Four inspectable contributions were delivered:

C1. Determinant-driven requirements specification. An eight-code implementation determinant taxonomy (D1–D8) was operationalized as requirements signals, enabling consistent translation from field evidence to deployable constraints. The resulting determinant concentrations reported in the Results section support the claim that first-use legibility and onboarding (D3), marker robustness and recovery (D2), and curriculum framing (D1) are primary adoption and enactment gates in city-scale deployments.

C2. Verifiable requirements catalogue (REQ-1 to REQ-18). Determinant signals were translated into a minimal requirements set using “shall” statements with verification cues, aligned with requirements engineering guidance for specifying testable requirements and quality attributes [5,42]. This approach prioritizes recoverability, portability in the context of BYOD, and safety-conscious operability as essential requirements rather than mere facilitative guidance.

C3. Evidence-to-requirement-to-artefact traceability. A bidirectional trace spine was provided linking determinants, requirements, transfer artefacts, and evidence anchors, addressing the classical requirements traceability problem and its auditability implications [12,55]. This design also responds to known gaps and barriers in pre-requirements traceability by privileging compact, maintainable trace structures over exhaustive linking [13,15].

C4. Operations-ready transfer kit and minimal operations stack. A minimal operations stack (roles, routines, maintenance procedures, incident response, and BYOD fallback) was specified as an implementation interface for replication. This packaging is in accordance with the anticipations of reproducibility in the realm of applied computing, wherein replication is contingent upon clearly defined artefacts, systematic procedures, and versioned assets, rather than relying solely on narrative exposition [41,56].

Collectively, the paper’s primary outcome is not a pedagogical effect claim. Robust implementation is treated as a prerequisite for interpretable learning evaluation in future work, consistent with contemporary process evaluation and fidelity guidance for complex interventions [63,65]. Instead, it is an auditable, determinant-driven engineering specification intended to reduce fragility in replication and support responsible implementation in public-space school contexts.

## 6.2. Limitations

This research reveals nine main limitations. These limitations delimit interpretation, validity, and transfer inferences.

First, the study is framed as a feasibility and traceability contribution for an in-the-wild mobile AR service, rather than as an evaluation of learning effectiveness. Accordingly, teachers and students’ indicators are interpreted as adoption-relevant constraints (for example, feasibility and acceptability), not as evidence of competence development [7].

Second, construct validity and determinant granularity: Determinants are operationalized as single-label codes applied to meaning units to improve coding consistency and support quantification. This choice can underrepresent multi-determinant interactions, for example when BYOD constraints co-occur with onboarding friction. Pairing meaning-unit counts with record coverage (Section 4.3) reduces, but does not eliminate, interpretive ambiguity and coder subjectivity [53].

Third, internal validity and derivation bias: Requirements derivation is grounded in the same evidence base used to quantify determinant concentrations. Although the derivation procedure is specified and auditable (Sections 3.8-3.9), analyst judgement can still influence how statements are

elevated into “shall” requirements. The traceability matrix mitigates this risk by forcing explicit evidence anchors per requirement and by aligning with established traceability practice and requirements-engineering guidance [5,12].

Fourth, external validity and context dependence: Evidence derives from a single city-scale deployment with a fixed set of POIs and tasks. While the determinant’s logic is expected to transfer to comparable in-the-wild AR services, local differences in urban morphology, governance arrangements, maintenance capacity, and school device policies can shift determinant priority and alter feasible mitigations [4].

Fifth, marker-centric interaction path: The path architecture retains a structural dependency on marker-triggered access at key stages, elevating the marker-related determinant as a gating constraint. Alternative anchoring and tracking approaches could reduce this dependency, but such alternatives were not evaluated in the present deployment [2].

Sixth, instrumentation validity and unit of analysis: Telemetry is treated strictly as a feasibility descriptor, but logging completeness and interpretation remain dependent on event schemas, device behavior, and field conditions. In addition, the unit of analysis is the group session rather than the individual learner. This is consistent with the service boundary adopted, but it limits interpretability for user-level behavior modelling and performance questions.

Seventh, outcome validity and avoidance of overclaim: Post-path indicators and narrative evidence may reflect feasibility, perceived relevance, and adoption constraints, not competence development. This boundary is maintained deliberately, given variability in AR learning studies and the frequent reliance on limited pre-post evidence in the broader literature [68].

Eighth, traceability maintenance cost is not measured: While traceability is structured to reduce overhead, the operational cost of maintaining trace links across iterative content updates and software releases was not empirically measured. This is still a practical risk, consistent with reported barriers to traceability adoption in practice [15].

Ninth, privacy and governance constraints limit broader release: Data minimization and GDPR-aligned governance restrict the degree of open release possible for some operational artefacts and telemetry, requiring balancing transparency with lawful processing and proportionality [17,39].

The limitations constrain inference but inspire a synthesis of contributions and future paths.

### 6.3. Future Paths

Several concrete research and engineering paths follow directly from the determinant concentrations, requirements catalogue, and operations stack.

Cross-city replication with controlled transfer evaluation. Replication studies across municipalities should measure adoption latency, enactment breakdown rates, and operational workload when the transfer kit is used by non-originating teams. This would directly test whether determinant-driven artefacts reduce fragility in practice [1,13].

Recovery instrumentation as first-class telemetry. Future versions of the logging schema should explicitly encode recovery events, such as recognition failure, rescan attempts, fallback activation, and regrouping interruptions. This would enable more precise validation of D2 and D3 requirements under field conditions without shifting the paper’s boundary into learning outcomes.

Onboarding friction experiments and legibility benchmarking. Controlled studies should compare onboarding variants (quick-start formats, role cards, scanning guidance) and quantify their impact on first-use errors, time-to-first-successful trigger, and resumption efficiency, aligning with usability-in-context principles [5].

Automated marker health monitoring and maintenance optimization. Operational tooling should be developed for marker health inspection, including periodic audits, glare risk scoring by placement, and replacement scheduling. This would operationalize D2 as a maintainability and reliability problem rather than a manual facilitation burden [42].

Tool support for traceability maintenance. Lightweight tooling should support semi-automated updates to the determinant-to-requirement-to-artefact links when content is revised. This addresses

a key adoption barrier identified in traceability practice studies, namely the perceived cost of maintaining trace links [15,55].

Privacy-preserving analytics for operations. Privacy-aware telemetry designs should be evaluated, including aggregation strategies, short retention windows, and risk-based governance protocols, to enable operational auditing while preserving minimization and lawful processing principles [17,39].

Reduced dependency on brittle triggers. Alternative progression mechanisms should be explored for contexts where marker deployment is infeasible or where smartphone restrictions apply, expanding the D8 fallback space beyond procedural mitigations and into architectural choices [3].

**Author Contributions:** Conceptualization, J.F.-S.; methodology, J.F.-S.; validation, J.F.-S. and L.P.; formal analysis, J.F.-S.; investigation, J.F.-S.; resources, J.F.-S.; data curation, J.F.-S.; Writing—Original draft, J.F.-S.; Writing—Review and editing J.F.-S. and L.P.; visualization, J.F.-S.; supervision, L.P.; project administration, J.F.-S. and L.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by National Funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P., under Grant Number 2023.00257.BD., with the following DOI: <https://doi.org/10.54499/2023.00257.BD>. The EduCITY project is funded by National Funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project PTDC/CED-EDG/0197/2021.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, approved by the GDPR (27 November 2024), and approved by the Ethics Committee of University of Aveiro (protocol code 1-CE/2025 on 5 February 2025).

**Informed Consent Statement:** Informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The datasets supporting the findings of this study were generated during the implementation of the *Art Nouveau Path* mobile augmented reality game in Aveiro, Portugal. The raw research datasets (student questionnaires S1-PRE, S2-POST, and S3-FU; teacher reflection forms T1-R; and teacher observation records T2-OBS) are not publicly available due to GDPR and ethical restrictions. Versions of these datasets may be made available by the corresponding authors upon reasonable request, subject to institutional approval and applicable data-sharing conditions. To support transparency, non-sensitive instruments and aggregated resources are openly available in the project's Zenodo community "*Art Nouveau Path*", including: the complete *Art Nouveau Path* MARG and its mapping to the GreenComp framework (DOI: <https://doi.org/10.5281/zenodo.16981236>), and the automated gameplay logs summary (DOI: <https://doi.org/10.5281/zenodo.17507328>). All publicly shared files omit sensitive fields, and full item-level gameplay logs are available upon reasonable request under the same ethical and institutional conditions.

**Acknowledgments:** The authors acknowledge the support of the research team of the EduCITY project. The authors also appreciate the willingness of the participants to contribute to this study. During the preparation of this manuscript, the authors used Microsoft Word, Excel and PowerPoint (Microsoft 365), DeepL (DeepL Free Translator) was used to translate selected passages from Portuguese to English, ChatGPT (GPT-5, released 7 August 2025), R (version 4.4.1) and Julius.AI for the respective purposes of writing and editing text, cleaning and organizing data, designing schemes and tables, translation and language improvement, statistical analysis and data visualization, and cross checking descriptive statistics, clustering procedures and wording consistency. All outputs were treated as suggestions. Quantitative data were initially cleaned and preprocessed in Excel and subsequently analyzed and visualized in R (version 4.4.1) using the *tidyverse* ecosystem and *ggplot2* to generate publication quality figures. The authors have reviewed and edited all outputs and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
BYOD	Bring Your Own Device
POI	Point of Interest
T1-VAL	Teachers' Validation Questionnaire (Workshop)
T2-OBS	Teachers' Observation Questionnaire
S2-POST	Student's Post Questionnaire
HCI	Human-Computer Interaction
OS	Operating System
ISO	International Organization for Standardization
ML	Mobile Learning
NFR	Non-Functional Requirement
MARG	Mobile Augmented Reality Game
C1-C4	Contribution
RQ	Research Question
NIST	National Institute of Standards and Technology
ACM	Association for Computing Machinery
GDPR	General Data Protection Regulation
D1-D8	Determinant
REQ1-REQ18	Requirements
A1-A6	Artefacts
E_ID	Evidence
E_LOG	Logs (Gameplay)
GPS	Global Positioning System
FAIR	Findable, Accessible, Interoperable, Reusable
GCQuest	GreenComp-Based Questionnaire
$\alpha$	Alpha
OBS	Observation
KNOW	Knowledge
MDN	Median
M	Mean
IQR	Interquartile Range
SD	Standard Deviation
MU	Meaning Unit
FR	Functional Requirement
OP	Operational Requirement
DTLE	Digital Teaching and Learning Ecosystem
RACI	Responsible, Accountable, Consulted, Informed

## Appendix A. Full Determinant-to-Transfer Traceability Matrix

**Table A1.** Determinant-to-transfer kit traceability matrix (absolute counts and component lists; MU N = 131; teacher records N = 54).

Determinant (absolute Evidence)	Determinants						A6 Operations runbook pack (including differentiation and accessibility)
	A1 Preparation and legibility	A2 Curriculum framing	A3 Path orchestration and safety	A4 Technical robustness and fallback	A5 Consolidation and follow-up		

<b>D1</b> <i>Curriculum alignment and framing</i>	MU n = 24; Teachers' records n = 21/54	n/a	Curriculum mapping matrix; facilitation and framing script	n/a	n/a	n/a	n/a
<b>D2</b> <i>Marker robustness and recovery</i>	MU n = 25; Teachers' records n = 22/54	n/a	n/a	n/a	Marker deployment guidance; recovery steps; alternative triggers	n/a	n/a
<b>D3</b> <i>Usability and design clarity</i>	MU n = 29; Teachers' records n = 25/54	Teacher-facing quick start; onboarding notes; in-app legibility supports	n/a	n/a	n/a	n/a	n/a
<b>D4</b> <i>Post-activity consolidation and follow-up</i>	MU n = 14; Teachers' records n = 13/54	n/a	n/a	n/a	n/a	Structured debrief template; classroom follow-up prompts	n/a
<b>D5</b> <i>Differentiation and accessibility</i>	MU n = 14; Teachers' records n = 12/54	n/a	n/a	n/a	n/a	n/a	Age variants; accessibility notes
<b>D6</b> <i>Safety and supervision</i>	MU n = 8; Teachers' records n = 7/54	n/a	n/a	Safety briefing; supervision cues; regroup scripts	n/a	n/a	n/a
<b>D7</b>	MU n = 7;	n/a	n/a	Role cards; device-	n/a	n/a	n/a

<i>Collaboration and accountability routines</i>	Teachers' records n = 6/54				sharing protocol			
<i>D8 BYOD heterogeneity and low-tech fallback</i>	MU n = 10; Teachers' records n = 9/54	n/a	n/a	n/a	Compatibility checks; device prep; low-tech fallback	n/a	n/a	n/a

## Appendix B. Operational Templates and Runbook (Transfer-Ready)

Table A2. Orchestration checklist excerpts (transfer-facing).

Checklist area	Purpose	Operational cue
<i>Pacing buffers</i>	Allocate time buffers for interruptions	Use a buffer of 5 to 10 minutes for crossings and regrouping
<i>Technical contingencies</i>	Ensure recovery paths if AR or connectivity fails	Provide fallback prompts and non-AR progression cues
<i>Group management</i>	Maintain visibility and accountability in public space	Use headcounts and role rotation at each POI
<i>Start and end routines</i>	Reduce friction at launch and closure	Standardize a start script and debrief closure questions

Table A3. Transfer kit artefact inventory (A1 to A6).

Artefact	Name	Scope (what it contains)
A1	Preparation and legibility pack	Quick-start; onboarding notes; outdoor legibility checklist
A2	Curriculum framing pack	Curriculum-to-task mapping matrix; facilitation and framing script
A3	Path orchestration and safety pack	Safety briefing; crossing and regroup routines; role cards; device-sharing guidance
A4	Technical robustness and fallback pack	Marker production and placement guidance; recovery protocol; alternative triggers; BYOD checklist; offline or no-phone alternative
A5	Consolidation and follow-up pack	Debrief template; classroom follow-up prompts
A6	Operations runbook pack	Roles, routines, maintenance cycle, incident response, data handling and minimization guidance; adaptation variants by age and ability; accessibility notes.

Table A4. Minimal RACI for operations-ready replication.

Activity	Instructional lead	Technical steward	Content maintainer
----------	--------------------	-------------------	--------------------

<i>Curriculum framing and teacher briefing</i>	R/A	C	C
<i>BYOD readiness and device triage</i>	C	R/A	C
<i>Marker deployment and inspection</i>	C	C	R/A
<i>In-session recovery and fallback activation</i>	R	C	C
<i>Post-session log retrieval and upload</i>	R	C	C
<i>Incident logging and corrective actions</i>	R	C	A
<i>Periodic marker maintenance cycle</i>	C	C	R/A

Note: R = responsible; A = accountable; C = consulted.

**Table A5.** Minimal operations routines (operations-ready view).

<b>Routine</b>	<b>Objective</b>	<b>Primary role(s)</b>	<b>Inputs</b>	<b>Outputs</b>	<b>Frequency</b>
<i>Pre-Session Preparation</i>	Reduce first-use friction; manage BYOD heterogeneity	Instructional lead; Technical steward	A1; A4	Devices checked; markers inspected; groups formed	Per session
<i>Session Launch</i>	Standardize onboarding and safety briefing	Instructional lead	A2; A3	Roles assigned; pacing buffers announced; session started	Per session
<i>POI Enactment Loop</i>	Maintain pacing and accountability at POIs	Instructional lead	A3; app tasks	POI blocks completed; regroup checks executed	Per POI
<i>Interruption Recovery</i>	Sustain continuity under recognition or connectivity failure	Instructional lead; Technical steward or (if present)	A4	Session continues via recovery or fallback	As needed
<i>Path Closure and Debrief</i>	Consolidate and close activity	Instructional lead	A5	Debrief captured; follow-up prompts assigned	Per session
<i>Post-session log transfer</i>	Preserve feasibility evidence; support audit	Instructional lead; Content maintainer	A6	Logs uploaded; integrity checks performed	Per session
<i>Marker maintenance cycle</i>	Sustain marker health in city environment	Content maintainer	A4; A6	Markers replaced or re-mounted; issues logged	Weekly or monthly

## Appendix C. Logging Schema and Redacted Example Record

**Table A6.** Minimal group-session logging schema (feasibility-only usage).

Field	Type	Description	Notes (privacy/usage in this paper)
<i>session_id</i>	string	Anonymous group-session identifier	No direct personal identifiers; used to aggregate events per session.
<i>timestamp</i>	datetime	Event timestamp (device-local or normalized)	Resolution sufficient for duration envelopes; do not infer individual behavior.
<i>group_size</i>	integer	Approximate number of students in the group	Optional; if unavailable, store null.
<i>poi_id</i>	string	Point-of-interest identifier (P1 to P8)	Maps events to POI-level completion traces.
<i>task_id</i>	string	Task identifier within POI (T01 to T36)	Supports task-level completion presence only.
<i>stage_id</i>	string	Stage identifier in the path	Used for progression and resumption analysis.
<i>access_mode</i>	categorical	Stage entry trigger used (e.g., ARBook, AR marker)	Represents marker-mediated access modality.
<i>event_type</i>	categorical	Event type (e.g., stage_start, item_presented, response_submitted, stage_end, interruption, resume)	Recommended to extend with explicit recovery events in future work.
<i>response_present</i>	boolean	Whether a response was recorded for the presented item	Used as feasibility indicator; correctness not used in this paper.
<i>duration_s</i>	number	Elapsed time for task or stage (seconds)	Used to compute session duration envelopes only.
<i>device_os</i>	categorical	Client OS (Android/iOS) and version	Used to characterize BYOD heterogeneity.
<i>app_version</i>	string	Mobile client version	Supports replication and version control.

**Table A7.** Redacted example log record (illustrative; no personal identifiers).

Field	Value (redacted example)
<i>session_id</i>	S_2025_05_17_013
<i>timestamp</i>	2025-05-17T10:42:31Z
<i>group_size</i>	4
<i>poi_id</i>	P3
<i>task_id</i>	T14

<i>stage_id</i>	S3
<i>access_mode</i>	AR marker
<i>event_type</i>	response_submitted
<i>response_present</i>	TRUE
<i>duration_s</i>	47
<i>device_os</i>	Android 14
<i>app_version</i>	1.3

## Appendix D

This appendix operationalizes REQ to E to D to A traceability using instantiated evidence anchors. Evidence anchors (E\_ID) are minimal, audit-ready pointers to a source record identifier or to a log pattern that supports inspection.

Anchor inventory (canonical counts): T1-VAL teacher records n = 30; T2-OBS teacher records n = 24; LOGS session records n = 118. Total instantiated evidence anchors n = 172.

### Appendix D.1. Instantiated Evidence Anchors (E\_ID).

**Table A8.** Evidence anchors (T1-VAL teacher records).

E_ID	Teacher	Subject
<i>E-T1VAL-R001</i>	Teacher_1	Arts
<i>E-T1VAL-R002</i>	Teacher_2	Geography
<i>E-T1VAL-R003</i>	Teacher_3	Multidisciplinary
<i>E-T1VAL-R004</i>	Teacher_4	Mathematics
<i>E-T1VAL-R005</i>	Teacher_5	Geography
<i>E-T1VAL-R006</i>	Teacher_6	Science
<i>E-T1VAL-R007</i>	Teacher_7	Mathematics
<i>E-T1VAL-R008</i>	Teacher_8	Civic Education
<i>E-T1VAL-R009</i>	Teacher_9	Multidisciplinary
<i>E-T1VAL-R010</i>	Teacher_10	Arts
<i>E-T1VAL-R011</i>	Teacher_11	Civic Education
<i>E-T1VAL-R012</i>	Teacher_12	Mathematics
<i>E-T1VAL-R013</i>	Teacher_13	History
<i>E-T1VAL-R014</i>	Teacher_14	History
<i>E-T1VAL-R015</i>	Teacher_15	Arts
<i>E-T1VAL-R016</i>	Teacher_16	Arts
<i>E-T1VAL-R017</i>	Teacher_17	Civic Education
<i>E-T1VAL-R018</i>	Teacher_18	Science
<i>E-T1VAL-R019</i>	Teacher_19	Civic Education
<i>E-T1VAL-R020</i>	Teacher_20	Multidisciplinary
<i>E-T1VAL-R021</i>	Teacher_21	Civic Education
<i>E-T1VAL-R022</i>	Teacher_22	Arts
<i>E-T1VAL-R023</i>	Teacher_23	Geography
<i>E-T1VAL-R024</i>	Teacher_24	Multidisciplinary
<i>E-T1VAL-R025</i>	Teacher_25	Geography

<i>E-T1VAL-R026</i>	Teacher_26	History
<i>E-T1VAL-R027</i>	Teacher_27	Geography
<i>E-T1VAL-R028</i>	Teacher_28	Civic Education
<i>E-T1VAL-R029</i>	Teacher_29	Mathematics
<i>E-T1VAL-R030</i>	Teacher_30	Mathematics

**Table A9.** Evidence anchors (T2-OBS teacher records).

<b>E_ID</b>	<b>Teacher</b>
<i>E-T2OBS-R001</i>	Teacher_1
<i>E-T2OBS-R002</i>	Teacher_2
...	...
<i>E-T2OBS-R023</i>	Teacher_23
<i>E-T2OBS-R024</i>	Teacher_24

**Table A10.** Evidence anchors (LOGS session records). SheetRow refers to the row in the LOGS sheet excluding the header row.

<b>E_ID</b>	<b>SheetRow</b>
<i>E-LOG-R001</i>	1
<i>E-LOG-R002</i>	2
<i>E-LOG-R003</i>	3
...	...
<i>E-LOG-R117</i>	117
<i>E-LOG-R118</i>	118

*Appendix D.2. REQ to E to D to A Mapping*

In Table D4, the E column lists the eligible evidence pools used to justify each requirement. Point-wise anchors per requirement can be derived once MU-level coding or requirement-specific tagging is available.

**Table A11.** REQ towards E towards D and towards A traceability (evidence pools).

<b>REQ</b>	<b>Evidence anchors (E)</b>	<b>D</b>	<b>A</b>
<i>REQ-01</i>	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D1	A2
<i>REQ-02</i>	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D1	A2
<i>REQ-03</i>	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D3	A1
<i>REQ-04</i>	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D3	A1
<i>REQ-05</i>	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D3	A1
<i>REQ-06</i>	T2-OBS: E-T2OBS-R001 to E-T2OBS-R024; LOGS: E-LOG-R001 to E-LOG-R118	D2	A4

REQ-07	T2-OBS: E-T2OBS-R001 to E-T2OBS-R024; LOGS: E-LOG-R001 to E-LOG-R118	D2	A4
REQ-08	T2-OBS: E-T2OBS-R001 to E-T2OBS-R024; LOGS: E-LOG-R001 to E-LOG-R118	D2	A4
REQ-09	T2-OBS: E-T2OBS-R001 to E-T2OBS-R024; LOGS: E-LOG-R001 to E-LOG-R118	D8	A4
REQ-10	T2-OBS: E-T2OBS-R001 to E-T2OBS-R024; LOGS: E-LOG-R001 to E-LOG-R118	D8	A4
REQ-11	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D6	A3
REQ-12	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D6	A3
REQ-13	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D7	A3
REQ-14	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D7	A3
REQ-15	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D4	A5
REQ-16	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D4	A5
REQ-17	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D5	A6
REQ-18	T1-VAL: E-T1VAL-R001 to E-T1VAL-R030; T2-OBS: E-T2OBS-R001 to E-T2OBS-R024	D5	A6

## References

1. Rogers, Y. Interaction Design Gone Wild. *Interactions* **2011**, *18*, 58–62, doi:10.1145/1978822.1978834.
2. Syed, T.A.; Siddiqui, M.S.; Abdullah, H.B.; Jan, S.; Namoun, A.; Alzahrani, A.; Nadeem, A.; Alkhodre, A.B. In-Depth Review of Augmented Reality: Tracking Technologies, Development Tools, AR Displays, Collaborative AR, and Security Concerns. *Sensors* **2022**, *23*, 146, doi:10.3390/s23010146.
3. Howell, G.; Franklin, J.M.; Sritapan, V.; Souppaya, M.; Scarfone, K. *Guidelines for Managing the Security of 34 Mobile Devices in the Enterprise*; 2023;
4. Stefanidi, H.; Sünderkamp, J.-H.; Tatzgern, M.; Itzlinger, A.; Meschtscherjakov, A. You're Making Things AR-Kward: Exploring Augmented Reality In-the-Wild MHCI041. *Proc. ACM Human-Computer Interact.* **2025**, *9*, 1–24, doi:10.1145/3743740.
5. ISO/IEC/IEEE. ISO/IEC/IEEE 29148:2018 Systems and software engineering Life cycle processes Requirements engineering; 2nd ed.; International Organization for Standardization: Geneva, Switzerland, 2018. Available online: <https://www.iso.org/standard/72089.html> (accessed on 24 February 2026).
6. Kim, J.S. Making Every Study Count: Learning From Replication Failure to Improve Intervention Research. *Educ. Res.* **2019**, *48*, 599–607, doi:10.3102/0013189X19891428.
7. Proctor, E.; Silmere, H.; Raghavan, R.; Hovmand, P.; Aarons, G.; Bunger, A.; Griffey, R.; Hensley, M. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Adm. Policy Ment. Heal. Ment. Heal. Serv. Res.* **2011**, *38*, 65–76, doi:10.1007/s10488-010-0319-7.
8. Liang, L.; Zhang, Z.; Guo, J. The Effectiveness of Augmented Reality in Physical Sustainable Education on Learning Behaviour and Motivation. *Sustainability* **2023**, *15*, 5062, doi:10.3390/su15065062.

9. Beyer, B.; Jones, C.; Petoff, J.; Murphy, N.R. *Site Reliability Engineering: How Google Runs Production Systems*; Beyer, B., Jones, C., Petoff, J., Murphy, N.R., Eds.; O'Reilly Media, Inc., 2016; ISBN 978-1-491-92912-4.
10. ISO/IEC. ISO/IEC 25010:2023 Systems and software engineering Systems and software Quality Requirements and Evaluation (SQuaRE) Product quality model; 2nd ed.; International Organization for Standardization: Geneva, Switzerland, 2023. Available online: <https://www.iso.org/standard/78176.html> (accessed on 24 February 2026).
11. Krause, J.; Kaufmann, A.; Riehle, D. The Code System of a Systematic Literature Review on Pre-Requirements Specification Traceability; 2020;
12. Gotel, O.C.Z.; Finkelstein, C.W. An Analysis of the Requirements Traceability Problem. In Proceedings of the Proceedings of IEEE International Conference on Requirements Engineering; IEEE Comput. Soc. Press; pp. 94–101.
13. Mucha, J.; Kaufmann, A.; Riehle, D. A Systematic Literature Review of Pre-Requirements Specification Traceability. *Requir. Eng.* **2024**, *29*, 119–141, doi:10.1007/s00766-023-00412-z.
14. Moran, K.; Palacio, D.N.; Bernal-Cárdenas, C.; McCrystal, D.; Poshyvanyk, D.; Shenefiel, C.; Johnson, J. Improving the Effectiveness of Traceability Link Recovery Using Hierarchical Bayesian Networks. In Proceedings of the Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering; ACM: New York, NY, USA, June 27 2020; pp. 873–885.
15. Ruiz, M.; Hu, J.Y.; Dalpiaz, F. Why Don't We Trace? A Study on the Barriers to Software Traceability in Practice. *Requir. Eng.* **2023**, *28*, 619–637, doi:10.1007/s00766-023-00408-9.
16. Lundmark, R.; Hasson, H.; Richter, A.; Khachatryan, E.; Åkesson, A.; Eriksson, L. Alignment in Implementation of Evidence-Based Interventions: A Scoping Review. *Implement. Sci.* **2021**, *16*, 93, doi:10.1186/s13012-021-01160-w.
17. Boeckl, K.R.; Lefkowitz, N.B. NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management, Version 1.0; 2020;
18. Houben, S.; Marquardt, N.; Vermeulen, J.; Schöning, J.; Klokose, C.; Reiterer, H.; Korsgaard, H.; Schreiner, M. Cross-Surface: Challenges and Opportunities for “bring Your Own Device” in the Wild. In Proceedings of the Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; ACM: New York, NY, USA, May 7 2016; pp. 3366–3372.
19. Hoerber, O.; Harvey, M.; Dewan Sagar, S.A.; Pointon, M. The Effects of Simulated Interruptions on Mobile Search Tasks. *J. Assoc. Inf. Sci. Technol.* **2022**, *73*, 777–796, doi:10.1002/asi.24579.
20. Schneegass, C.; Füseschi, V.; Konevych, V.; Draxler, F. Investigating the Use of Task Resumption Cues to Support Learning in Interruption-Prone Environments. *Multimodal Technol. Interact.* **2021**, *6*, 2, doi:10.3390/mti6010002.
21. ISO. ISO 9241-11:2018 Ergonomics of human-system interaction Part 11: Usability: Definitions and concepts; 2nd ed.; International Organization for Standardization: Geneva, Switzerland, 2018. Available online: <https://www.iso.org/standard/63500.html> (accessed on 24 February 2026).
22. Weiser, M. The Computer for the 21st Century. *Sci. Am.* **1991**, *265*, 94–104, doi:10.1038/scientificamerican0991-94.
23. Yoon, H.; Shin, C. Cross-Device Computation Coordination for Mobile Collocated Interactions with Wearables. *Sensors* **2019**, *19*, 796, doi:10.3390/s19040796.
24. Gotel, O.; Cleland-Huang, J.; Hayes, J.H.; Zisman, A.; Egyed, A.; Grünbacher, P.; Dekhtyar, A.; Antoniol, G.; Maletic, J.; Mäder, P. Traceability Fundamentals. In *Software and Systems Traceability*; Springer London: London, 2012; pp. 3–22.
25. Association for Computing Machinery. *Artifact Review and Badging, Current*; Version 1.1; 24 August 2020. Available online: <https://www.acm.org/publications/policies/artifact-review-and-badging-current> (accessed on 24 February 2026).
26. Association for Computing Machinery. *New Changes to Badging Terminology*; Document last revised 24 August 2020. Available online: <https://www.acm.org/publications/badging-terms> (accessed on 24 February 2026).
27. Piccolo, S.R.; Frampton, M.B. Tools and Techniques for Computational Reproducibility. *Gigascience* **2016**, *5*, 30, doi:10.1186/s13742-016-0135-4.

28. Pham, Q.; Ton That, D.H.; Malik, T.; Youngdahl, A. Improving Reproducibility of Distributed Computational Experiments. In Proceedings of the Proceedings of the 1st International Workshop on Practical Reproducible Evaluation of Computer Systems, P-RECS 2018; 2018.
29. Pauzi, Z.; Thind, R.; Capiluppi, A. Artifact Traceability in DevOps: An Industrial Experience Report. In Proceedings of the Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering; ACM: New York, NY, USA, June 14 2023; pp. 180–183.
30. Blanco-Pons, S.; Carrión-Ruiz, B.; Duong, M.; Chartrand, J.; Fai, S.; Lerma, J.L. Augmented Reality Markerless Multi-Image Outdoor Tracking System for the Historical Buildings on Parliament Hill. *Sustainability* **2019**, *11*, 4268, doi:10.3390/su11164268.
31. Panou, C.; Ragia, L.; Dimelli, D.; Mania, K. An Architecture for Mobile Outdoors Augmented Reality for Cultural Heritage. *ISPRS Int. J. Geo-Information* **2018**, *7*, 463, doi:10.3390/ijgi7120463.
32. Council of Europe Council Conclusions of 21 May 2014 on Cultural Heritage as a Strategic Resource for a Sustainable Europe. *Off. J. Eur. Union* **2014**, *57*, 36–38.
33. Bianchi, G.; Pisiotis, U.; Cabrera, M.; Punie, Y.; Bacigalupo, M. *The European Sustainability Competence Framework*; 2022; ISBN 9789276464853.
34. European Commission *Key Competences for Lifelong Learning*; Publications Office of the European Union; Luxembourg, 2019; ISBN 9789276004752.
35. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Digital Education Action Plan 2021-2027: Resetting Education and Training for the Digital Age; COM(2020) 624 final; European Commission: Brussels, Belgium, 30 September 2020. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0624> (accessed on 24 February 2026).
36. Redecker, C.; Punie, Y. *European Framework for the Digital Competence of Educators - DigCompEdu*; Luxembourg, 2017;
37. UNESCO *UNESCO Policy Guidelines for Mobile Learning*. Paris: UNESCO.; Kraut, R., Ed.; UNESCO: Paris, 2013; ISBN 9789230011437.
38. Vuorikari, R.; Kluzer, S.; Punie, Y. *DigComp 2.2, The Digital Competence Framework for Citizens - With New Examples of Knowledge, Skills and Attitudes*; Publications Office of the European Union: Luxembourg, 2022; ISBN 9789276488828.
39. European Parliament; Council of Europe Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation); European Parliament and Council of Europe: Brussels, 2016;
40. Peng, R.D. Reproducible Research in Computational Science. *Science (80)*. **2011**, *334*, 1226–1227, doi:10.1126/science.1213847.
41. Sandve, G.K.; Nekrutenko, A.; Taylor, J.; Hovig, E. Ten Simple Rules for Reproducible Computational Research. *PLoS Comput. Biol.* **2013**, *9*, e1003285, doi:10.1371/journal.pcbi.1003285.
42. ISO/IEC. ISO/IEC 25010:2011 Systems and Software Engineering: Systems and Software Quality Requirements and Evaluation (SQuaRE): System and Software Quality Models; International Organization for Standardization: Geneva, Switzerland, 2011. Available online: <https://www.iso.org/standard/35733.html> (accessed on 24 February 2026).
43. ISO/IEC. ISO/IEC 27001:2022 Information Security Management Systems: Requirements; International Organization for Standardization: Geneva, Switzerland, 2022. Available online: <https://www.iso.org/standard/27001> (accessed on 24 February 2026).
44. ISO/IEC. ISO/IEC 27002:2022 Information Security, Cybersecurity and Privacy Protection: Information Security Controls; International Organization for Standardization: Geneva, Switzerland, 2022. Available online: <https://www.iso.org/standard/75652.html> (accessed on 24 February 2026).
45. UNESCO *UNESCO Strategy in Technological Innovation in Education (2022-2025)*; Paris, 2021;
46. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018, doi:10.1038/sdata.2016.18.

47. ISO. ISO 31000:2018 Risk Management: Guidelines; International Organization for Standardization: Geneva, Switzerland, 2018. Available online: <https://www.iso.org/standard/65694.html> (accessed on 24 February 2026).
48. Hsieh, H.-F.; Shannon, S.E. Three Approaches to Qualitative Content Analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288, doi:10.1177/1049732305276687.
49. Fereday, J.; Muir-Cochrane, E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *Int. J. Qual. Methods* **2006**, *5*, 80–92, doi:10.1177/160940690600500107.
50. Miles, M.B.; Huberman, A.M.; Saldaña, J. *Qualitative Data Analysis: A Methods Sourcebook*; 4th ed.; SAGE Publications, Inc.: Los Angeles, 2020; ISBN 9781506353074.
51. Saldaña, J. *The Coding Manual for Qualitative Researchers*; 4th ed.; SAGE Publications Ltd., 2021; ISBN 978-1529731743.
52. Hayes, A.F.; Krippendorff, K. Answering the Call for a Standard Reliability Measure for Coding Data. *Commun. Methods Meas.* **2007**, *1*, 77–89, doi:10.1080/19312450709336664.
53. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*; SAGE Publications, Inc.: 2455 Teller Road, Thousand Oaks California 91320, 2019; ISBN 9781506395661.
54. GreenComp-Based Questionnaire (GCQuest) – ENG. Available online: <https://zenodo.org/records/14524933> (accessed on 24 February 2026).
55. Cleland-Huang, J.; Gotel, O.C.Z.; Huffman Hayes, J.; Mäder, P.; Zisman, A. Software Traceability: Trends and Future Directions. In Proceedings of the Future of Software Engineering Proceedings; ACM: New York, NY, USA, May 31 2014; pp. 55–69.
56. Nuseibeh, B.; Easterbrook, S. Requirements Engineering. In Proceedings of the Proceedings of the Conference on The Future of Software Engineering; ACM: New York, NY, USA, May 2000; pp. 35–46.
57. ISO/IEC. ISO/IEC 27701:2019 Security Techniques: Extension to ISO/IEC 27001 and ISO/IEC 27002 for Privacy Information Management: Requirements and Guidelines; International Organization for Standardization: Geneva, Switzerland, 2019. Available online: <https://www.iso.org/standard/71670.html> (accessed on 24 February 2026).
58. Pombo, L.; Marques, M.M. EduCITY as a Smart Learning City Environment towards Education for Sustainability - Work in Progress. *Proc. EdMedia + Innov. Learn.* **2023**, 133–139.
59. Ferreira-Santos, J.; Pombo, L. The Art Nouveau Path: Promoting Sustainability Competences Through a Mobile Augmented Reality Game. *Multimodal Technol. Interact.* **2025**, *9*, 77, doi:10.3390/mti9080077.
60. Ferreira-Santos, J.; Pombo, L. The Art Nouveau Path: Integrating Cultural Heritage into a Mobile Augmented Reality Game to Promote Sustainability Competences Within a Digital Learning Ecosystem. *Sustainability* **2025**, *17*, 8150, doi:10.3390/su17188150.
61. Ferreira-Santos, J.; Pombo, L. The Art Nouveau Path: From Gameplay Logs to Learning Analytics in a Mobile Augmented Reality Game for Sustainability Education. *Information* **2026**, *17*, 87, doi:10.3390/info17010087.
62. An, M.; Dusing, S.C.; Harbourne, R.T.; Sheridan, S.M. What Really Works in Intervention? Using Fidelity Measures to Support Optimal Outcomes. *Phys. Ther.* **2020**, *100*, 757–765, doi:10.1093/ptj/pzaa006.
63. Moore, G.F.; Audrey, S.; Barker, M.; Bond, L.; Bonell, C.; Hardeman, W.; Moore, L.; O’Cathain, A.; Tinati, T.; Wight, D.; et al. Process Evaluation of Complex Interventions: Medical Research Council Guidance. *BMJ* **2015**, *350*, h1258–h1258, doi:10.1136/bmj.h1258.
64. Rojas-Andrade, R.; Bahamondes, L.L. Is Implementation Fidelity Important? A Systematic Review on School-Based Mental Health Programs. *Contemp. Sch. Psychol.* **2019**, *23*, 339–350, doi:10.1007/s40688-018-0175-0.
65. Skivington, K.; Matthews, L.; Simpson, S.A.; Craig, P.; Baird, J.; Blazeby, J.M.; Boyd, K.A.; Craig, N.; French, D.P.; McIntosh, E.; et al. A New Framework for Developing and Evaluating Complex Interventions: Update of Medical Research Council Guidance. *BMJ* **2021**, n2061, doi:10.1136/bmj.n2061.
66. Pérez, D.; Van der Stuyft, P.; Zabala, M. del C.; Castro, M.; Lefèvre, P. A Modified Theoretical Framework to Assess Implementation Fidelity of Adaptive Public Health Interventions. *Implement. Sci.* **2015**, *11*, 91, doi:10.1186/s13012-016-0457-8.

67. Wiltsey Stirman, S.; Baumann, A.A.; Miller, C.J. The FRAME: An Expanded Framework for Reporting Adaptations and Modifications to Evidence-Based Interventions. *Implement. Sci.* **2019**, *14*, 58, doi:10.1186/s13012-019-0898-y.
68. Akçayır, M.; Akçayır, G. Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature. *Educ. Res. Rev.* **2017**, *20*, 1–11, doi:https://doi.org/10.1016/j.edurev.2016.11.002.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.