Check for updates

RESEARCH ARTICLE

Choosing what not to adapt: An instrument that can say no before traditional games are digitised for screen or hybrid play

Hatice Büber-Kaya[1], Meisam Taheri[2]

[1]Department of Software Engineering, Faculty of Engineering, Kırklareli University, Kırklareli, Kırklareli, 39100, Turkey
[2]Department of Game Development – The Game School, Faculty of Film, TV and Games, Inland Norway University of Applied Sciences - Hamar Campus, Hamar, Innlandet, 2317, Norway

v1 **First published:** 26 Nov 2026
https://doi.org/10.68336/pc4vvf09

Abstract

Background

Traditional games are increasingly proposed as source material for digital and hybrid adaptations, yet the decision of which documented games to adapt is usually informal. Frameworks exist for evaluating educational games and for formalising game rules, but we found none that screens a documented source game before adaptation and can return a reasoned non-selection decision. A five-criterion compensatory score applied to a 518-entry Turkish corpus placed 86% of candidates at 9 or 10 out of 10 and left two criteria constant.

Methods

We report an instrument whose output is a record, not a score: three dichotomous gates, an adaptation route assigned by test, five criteria, a transformation ledger stating what each rule element becomes on the target platform, and a separately logged distinctiveness field. A platform-conditional question (does the carrying mechanic survive under screen-only or sensor interfaces?) separates games suited to screen adaptation from hybrid candidates. A workbench uses a language model to draft a record per game with a source quotation per judgement; a human coder reviews each section.

Results

In a formative nine-game pilot selected for distinctiveness rather than to estimate corpus frequencies, the instrument produced reasoned non-selection decisions for five games; the earlier screening had vetoed all nine by keyword without case-specific reasons. Applying the instrument also exposed a non-discriminating gate in the instrument itself.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Inter-rater agreement varied substantially by field. Screen-survival classification showed high agreement (κ=.80), whereas the route field and all four scored criteria fell below the pre-specified.60 descriptor-revision threshold.

Conclusions

An instrument of this kind can return reasoned rejections before any adaptation exists. The pilot identifies five fields that need descriptor clarification before independent reliability evaluation. These formative agreement figures locate weaknesses in application consistency. They do not themselves establish or refute the validity of the selection inference.

Keywords
traditional games, game adaptation, selection instrument, rubric reliability, human-in-the-loop, large language models, cultural heritage, digital game design

This article is included in the COST Actions gateway.

Corresponding author: Hatice Büber-Kaya (hatice.buberkaya@klu.edu.tr)

Author roles: Büber-Kaya H: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Taheri M**: Formal Analysis, Methodology, Software, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action [CA22145], as supported by the COST Association (European Cooperation in Science and Technology).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Büber-Kaya H and Taheri M. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Büber-Kaya H and Taheri M. **Choosing what not to adapt: An instrument that can say no before traditional games are digitised for screen or hybrid play** Open Research Europe https://doi.org/10.68336/pc4vvf09

First published: 26 Nov 2026 https://doi.org/10.68336/pc4vvf09

1. Introduction

Digital technologies are increasingly considered alongside traditional and outdoor play, including attempts to carry traditional games into digital cultural contexts.[1,2] The proposals rarely state how the source game was chosen. In our own case, a preparatory screening of a 518-entry encyclopaedia of traditional games from Türkiye produced a shortlist of fifty games ranked by a five-criterion rubric. On inspection, the rubric could not reject anything, and this paper reports the instrument that replaced it.

The problem is pre-adaptation selection: given a game documented primarily as text, determine whether the available evidence supports its selection for a stated digital or hybrid interface, and record what is likely to be preserved or lost. Several literatures address parts of this problem. Ludemic approaches formalise traditional strategy-game rules through compositional elements.[3–5] They provide a basis for reasoning about rule decomposition. Their scope excludes games whose defining play depends on physical dexterity or bodily skill.[4] Work on exertion, pervasive and hybrid play identifies bodily, spatial and social properties that may be altered when play is mediated digitally.[6–10] This includes work explicitly concerned with traditional play or traditional sports.[8,10] Educational-game frameworks address design, analysis and evaluation from different directions.[11–14] They generally begin, however, from an educational-game problem in which learning goals or content are already part of the design brief. In the literature reviewed for this study we did not identify an instrument that combines these concerns into a pre-adaptation screen of a documented source game and can return a reasoned rejection.

A recent guide by Mesquita and Jucá[15] addresses the adjacent problem of adapting modern board games to digital environments. G.A.M.E. structures adaptation around a game's core elements, including mechanics, rules, narrative, aesthetics and technology, and was developed and refined as guidance for preserving that core across physical and digital versions. Our problem begins one step earlier: before an adaptation is designed, we ask whether the documented source game supports selection for a stated interface class at all. The present instrument also addresses traditional games whose game-defining competence may be bodily, spatial or socially enacted, so preservation cannot be assessed from formal rules alone. Its output is therefore not an adaptation guide but an auditable pre-adaptation record that can return a reasoned non-selection decision and state under which interface class the carrying mechanic survives.

We make three contributions. (1) A non-compensatory pre-adaptation selection instrument, centred on platform-conditional survival of the game's carrying mechanic. It outputs an auditable record rather than a total score. (2) A human-in-the-loop workbench that operationalises the instrument through structured model drafts and human adjudication. In the pilot, the human ratings were produced independently and compared with the model drafts afterwards, while the companion pipeline study (Taheri and Büber Kaya, unpublished report) reports the fuller implementation and evaluation. (3) Formative application cases. They show how the instrument produces traceable non-selection decisions and how applying the instrument itself exposed descriptor and gate defects that were subsequently revised.

2. Why the first rubric failed

The first screening applied three pass/fail gates (rules statable in two to three sentences; playable on a single mobile screen; adaptable to a literacy objective) and five criteria scored 0–2 (mechanic clarity MC, learning-objective mappability LOM, digitisability DIG, age-scalability AGE, cultural transferability CT), summed to a total out of ten. Of 518 entries, 163 were vetoed by keyword, 195 were read in full, 79 passed all gates and 50 were pooled. Table 1 shows what the scores did.

Thirteen games scored 10, thirty scored 9, and none scored below 7. The instrument as applied was total $= 4 + \mathrm{MC} + \mathrm{LOM} + \mathrm{CT}$: a three-criterion rubric with a fixed offset, reported as five criteria out of ten. Nine

Table 1. Score distribution of the first rubric over the 50-game pool.

Criterion	0	1	2	Behaviour
MC	0	9	41	at ceiling
LOM	0	27	23	only criterion that spread the pool (r with total $= 0.61$)
DIG	0	0	50	constant
AGE	0	0	50	constant
CT	1	10	39	weak ($r = 0.60$)

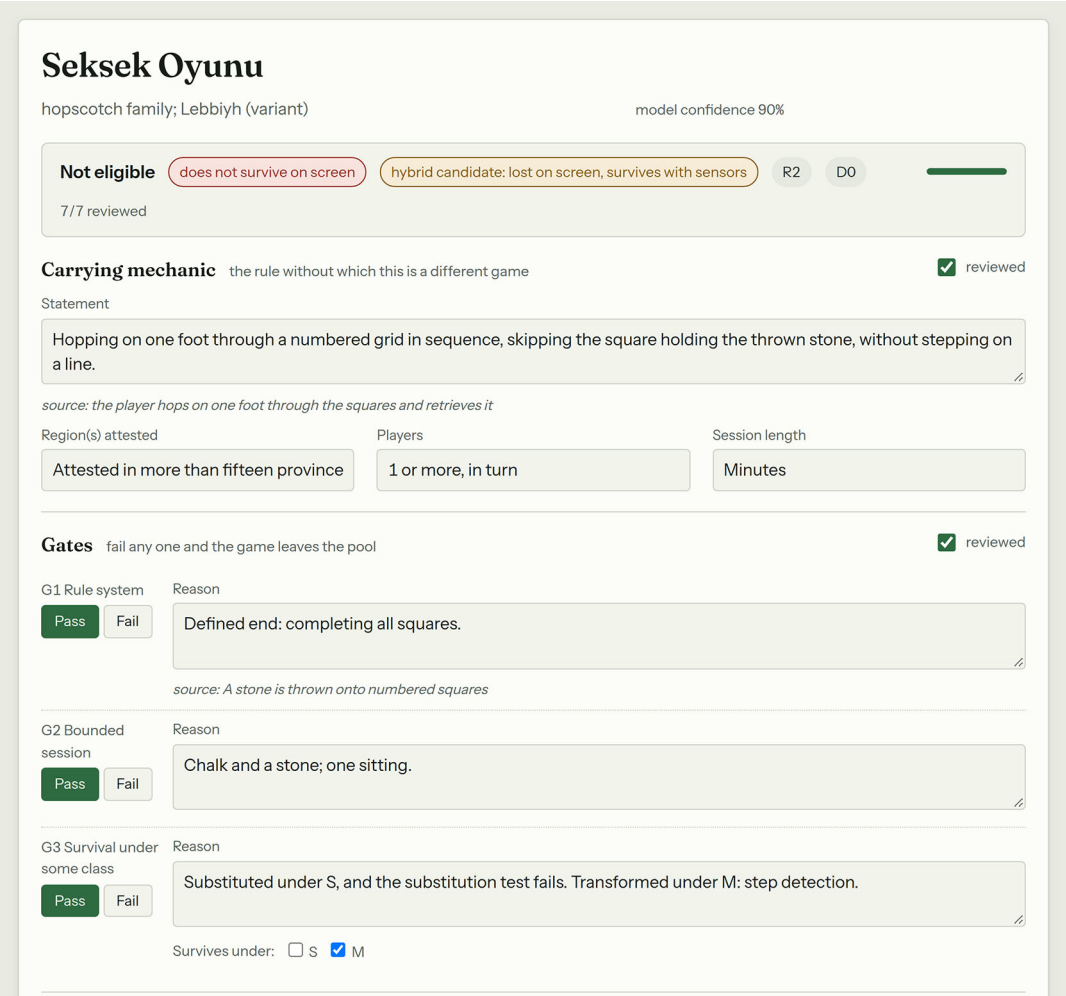


Figure 1. The record sheet for Seksek in the workbench. The verdict bar reads the eligibility rule off the record: not eligible on screen, hybrid candidate. A design illustration of the review mode; the pilot records were produced by independent coding (Section 6). Each judgement carries the source quotation the model paired with it, and each section is marked reviewed by a human.

target-throwing games (marbles, bottle caps, ring toss) averaged 9.0 while scoring the minimum on the only criterion concerned with learning.

Three faults explain this. Two criteria were consequences of the gates: anything that survives "playable on a single screen" is digitisable by construction, so DIG could not vary. Age-scalability was a property of the adaptation (vocabulary, timers) rather than of the game, so no game could score below 2. And LOM measured how easily letters could be attached to the game's objects, a test nothing can fail. For the present selection task, a rubric that cannot discriminate among candidates is not useful.

3. The instrument

The instrument is intended to support pre-adaptation selection of textually documented traditional games for digital or hybrid adaptation. It evaluates whether the documented evidence supports selection under a stated interface class. It does not predict the quality of a future implementation. Where an adaptation has an explicit educational objective, educational suitability is assessed additionally through LOM rather than treated as a universal condition of adaptation eligibility.

The revised instrument retains the five original dimensions conceptually but changes their operational form. MC, LOM, AGE and CT-a are scored criterion fields. CT-b and CT-c are descriptive fields. The former DIG dimension is represented

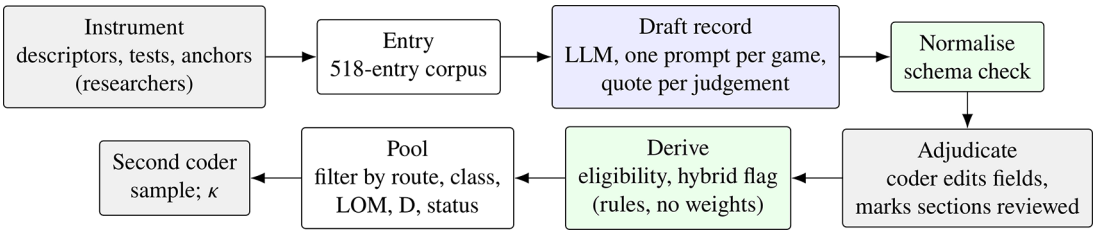


Figure 2. Pipeline. Grey: human steps. Blue: the model step. Green: deterministic rules. The instrument is written by the researchers before any extraction and its version is stored with each record.

by a platform-conditional survival profile rather than a score. The instrument adds structure before and after these fields and drops the total. A game passes through six parts in order (Figure 2). The output is a record.

3.1 Carrying mechanic

The minimal source-supported rule, action, or tightly coupled rule–action configuration whose removal would change the game-defining problem or turn the documented activity into a recognisably different game. The carrying mechanic states what must remain game-defining, and the survival test subsequently asks whether the competence required to enact that relation remains directly exercised, is transformed, or is acceptably substituted under a given interface. Coders identify it using three checks: the formulation must be anchored in the source entry, removing it must materially alter the game-defining problem, and no unnecessary component should remain in the formulation. Every later judgement addresses this element rather than the game's surface.

3.2 Gates

Three dichotomous gates address conditions that are conceptually distinct from the scored criteria. *G1 rule system:* the documented entry provides a sufficiently defined rule structure and end state to support the subsequent judgements. Failure of G1 is therefore a statement about the available documentation, not a claim that the traditional game itself lacks rules. *G2 implementation envelope:* one play forms a bounded session, and the documented setting, physical scale and required resources fall within what the intended mini-game format can feasibly accommodate. G2 concerns implementation scale and resource feasibility. G3 separately asks whether the carrying mechanic itself survives the target interface. *G3 survival:* the carrying mechanic *survives* under at least one of two interface classes: S (screen-only touch) or M (motion or sensor). A third class, L (co-located play with the device as referee or prop), is recorded but carries neither the gate nor the hybrid flag, for the reason given below. G3 operationalises a concern motivated by the movement-, exertion- and pervasive-game literature[6,9,16]: whether the game's carrying mechanic survives under a specified interface class. The first rubric asked nothing of the kind. Keyword vetoes such as "running" are replaced by a test each game takes individually.

3.3 A defect in G3 found by applying the instrument

The gate was originally stated as survival under *at least one* of S, M and L. Applying it to nine games, the coder recorded that the L class makes the gate non-discriminating: "if we put a camera or phone in front of any game it works with image processing … all games will pass, making this gate meaningless", and separately, "under L, almost every game comes out as *kept*, because a camera system suited to the game … can in principle act as the referee". DIG(L) was coded *kept* for eight of the nine games (Atlı Tura, coded *removed* on feasibility grounds, is the exception), including two judged impractical at their documented scale.

This is the objection of Section 2 applied to the present instrument, and we report it here together with its repair. If the device may sit beside a game played exactly as before, the carrying mechanic survives by definition and G3 cannot reject, in the same way that DIG could not vary under the first rubric. Two repairs were available: drop L from G3, or bound it to a sensing setup that must be realisable rather than merely conceivable. Both were proposed in the coder's own notes. The instrument now takes the first: **L is removed from G3** (v0.3), and a later revision (v0.4) settled a second ambiguity the same episode exposed: the gate's descriptor admitted only *kept* or *transformed*, while the eligibility rule treated a *substituted* class with a passing substitution test as surviving. G3 now reads that same survival predicate, so a game cannot be eligible on a class that does not carry its gate. Neither change moves a verdict on the nine pilot games. A game passes the gate only under S or M; L carries neither the gate nor the hybrid flag, which is computed from M alone. Because the physical game is unchanged under a co-located setup by definition, the only informative question about L is whether the proposed digital support is feasible, so L is recorded as a proposed setup with a feasibility judgement (yes/no/unclear) rather than as a survival operation. On the nine pilot games this changes no gate result and no verdict: every pass is already

Table 2. Level descriptors and anchor games for the four scored criteria. Anchors are from the 50-game pool.

	0	1	2
MC rule decomposability	A core rule cannot be operationalised from the documented source without inventing an unstated decision rule or resolving a consequential ambiguity outside the available evidence. *Eski Minder* (leader judges the frozen pose)	The rule structure can be decomposed, but one consequential step or competing formulation remains under-specified in the source. *Biliç* (riddle game and which-palm game under one name)	The core rule structure decomposes into distinguishable elements without requiring an unstated decision rule. *Üçtaş* (place, slide, align three)
LOM intrinsic objective	The stated target learning objective can be removed or replaced without materially changing the game-defining decisions, actions or winning strategy. (If no educational objective has been specified, LOM should be recorded as not assessed rather than as 0.) *Üçgen* (hitting the marble wins whatever is written on it)	The target objective affects player decisions or task demands, but the carrying mechanic and game-defining problem remain substantially intact if that educational content is removed or replaced. *Sekir* (memory of position persists without a syllable layer)	Performing the target learning behaviour is part of the carrying mechanic or necessary to solve the game-defining problem; removing it materially changes or collapses that problem. *Harf Oyunu* (word production is the move); *Mangala* (counting the sowing)
AGE developmental adaptability	The carrying mechanic depends on a relatively narrow developmental, bodily or cognitive competence, and the documented game provides no source-supported way to vary that demand without changing the mechanic. *El Kızartmaca* (reaction time)	The competence demand can be adjusted while preserving the carrying mechanic, but the adjustment would have to be introduced by the adaptation rather than being part of the documented game. *Kulaktan Kulağa* (phrase length)	The carrying mechanic is not narrowly development-bound, or the documented game already contains a rule-native parameter or variant that scales the competence demand while preserving the mechanic. *Üçtaş–Dokuztaş–Onikitaş* (3, 9, 12 pieces on one rule)
CT-a documented cultural legibility	A core rule or required practice cannot be interpreted confidently without culturally specific information absent from the documented evidence.	The game is interpretable with a bounded, source-supported cultural gloss. *Ay Dede* (rhyme must be rendered)	The core rules are interpretable from the documented evidence without additional cultural explanation. *Kulaktan Kulağa*

carried by S or M. Atlı Tura is the case that shows the repair was already implicit: its L was coded *removed* on feasibility and scale grounds, applying informally the practicability bound the second repair would have written into the descriptor. Both repairs came from the coder's own notes, and the finding came from a human applying the instrument to source entries (coding the nine games without reference to the model outputs) and writing down why a label did not fit. The model did not surface it: it and the coder disagreed on G3 in six of the nine games, always in the same direction (model fail, human pass), and in all six the model had flagged G3 as a low-confidence field. What this shows is that independent human coding, compared afterwards with the drafts, makes defects in the criteria visible. It does not show a benefit of the draft-checking interface, which was not used to produce these records, and it measures nothing about coding speed.

3.4 Adaptation route

The corpus separates into three architectures with different development costs, recorded rather than folded into a score. *R1 contest:* needs an opponent and produces a winner. *R2 closure:* closes without a winner and can be played alone. *R3 construction:* has a making layer in which the object the player builds determines how it performs. Routes describe architecture rather than suitability and do not enter eligibility. An ordered test assigns the primary route. The test assigns R3 construction first when the game contains a making phase whose output materially affects subsequent play or outcome. If R3 does not apply, the test assigns R1 contest when removing an opposing or structurally required other-player role eliminates the game-defining problem. If neither applies, the test assigns R2 closure when the game remains playable alone, reaches a defined end state and requires no opponent or structurally necessary second-player role. Where another route also applies, the record notes it as secondary.

3.5 Criteria with level descriptors

Each scored criterion has written descriptors for levels 0–2, consistent with descriptive-rubric guidance,[17] and an anchor game from the pool, a practice consistent with rubric-reliability guidance[18] (Table 2). Drawing on Habgood and Ainsworth's concept of intrinsic integration,[19] we operationalise LOM through a counterfactual test applied to a stated target learning objective: remove or replace the target content or behaviour and ask whether the game-defining decisions, strategy or problem materially change. The target objective and the gameplay behaviour through which it is enacted are recorded explicitly. We operationalise AGE as developmental adaptability, in line with developmental-fit principles in the science-of-learning literature[20]: the developmental competence required by the carrying mechanic and whether that demand can be varied without changing the mechanic. We distinguish scaling already present in the documented game from parameters introduced by an adaptation, and AGE is not a general judgement of age-appropriateness, safety or pedagogical suitability. CT is divided into three fields with different decision roles, reflecting cultural-heritage serious-game research and UNESCO's emphasis on community participation in safeguarding intangible heritage.[21,22] CT-a records documented cultural legibility and is thresholded because it asks whether the available evidence is sufficient to support interpretation of the game. CT-b records the degree to which culturally specific content is constitutive of the game, and CT-c records stated community, locality and provenance relations. CT-b and CT-c are recorded rather than scored because higher values would not represent better adaptation suitability.

3.6 Survival profile and transformation ledger

For each survival class S and M, the record states what happens to the human-adjudicated carrying mechanic. Co-located support (L) is recorded separately as a proposed setup with a feasibility judgement rather than as a carrying-mechanic survival operation. *Kept*: the player performs the same game-defining skill and the device changes nothing about how it is performed. *Transformed*: the same game-defining competence remains directly exercised, but the device senses, mediates or approximates the action differently. *Substituted*: the immediate player action or capability required by the carrying mechanic is replaced by a different capability. A substitution is not automatically accepted or rejected. The substitution test below determines its survival status. *Removed*: the element is gone. The skill judged is the one recorded as the carrying mechanic, not any other competence the game also exercises. Without that anchor, the label can be re-argued after the fact by naming a different competence. The former DIG dimension is represented by the platform-conditional survival profile rather than by an ordinal score. The S field records the carrying-mechanic operation under screen-only adaptation. It is read off a ledger with one row per rule or material element, using a five-operation vocabulary defined for this instrument (kept/transformed /substituted /removed/added), drawing on work on ludemes, analog-to-digital adaptation, and the mechanics/mechanisms distinction.[15,23,24] The ledger makes the selection judgement auditable at the level of individual rule, material and social elements, and it provides structured input for downstream adaptation analysis.

3.7 The substitution test

Substitution was originally an automatic rejection: a different skill under the old name was treated as a different game. Applying the instrument to nine games showed that this makes the field two-valued in practice and puts the whole rejecting burden on *removed*, which a coder had no occasion to use. A screen can depict almost any formal rule, so nothing was ever removed outright. Read that way, DIG(S) reduces to "can a screen representation be invented", to which the answer is always yes, and the criterion stops discriminating in the same manner as the first rubric's DIG.

The instrument therefore treats substitution as conditional. A substitution counts as survival only when both conditions hold: (1) the player still confronts the same game-defining problem, and (2) a game-defining competence embodied in the human-adjudicated carrying mechanic remains meaningfully exercised rather than merely represented by the substitute. The adjudicator records which competence is preserved and which is lost. Both conditions are required. Preserving a high-level objective alone (attack, escape, reach a goal) is insufficient, because almost any simulation can preserve such an objective while replacing the competence through which the game is played.

The anchors are drawn from the pilot. *Accept*: Araya Aldırmaca on a touch screen, where physical agility is lost but pincer positioning, deception, route choice and the enclose-before-safety problem all survive. *Reject*: Birdir Bir on a touch screen, where jumping, bodily clearance, balance and landing are replaced by tapping and do not transfer. Of the five games the coder marked substituted on screen, one was accepted and four rejected, so the criterion continues to reject while no longer rejecting by construction.

3.8 Distinctiveness

Distinctiveness is recorded separately from eligibility as one of three search-outcome categories. D0: a close counterpart is identified with no material difference in the carrying mechanic or other core rule relations. D1: a related family counterpart is identified but at least one carrying-mechanic or core-rule relation differs. D2: no close counterpart is

identified under the documented search protocol. These categories are not added to a score, and D2 is not a claim of uniqueness. In the first screening's pool, legibility and distinctiveness appeared close to inverses (the games with the strongest external precedent are the most legible), so they must not share a cell. From instrument v0.5 the D field remains unset until retrieval is complete and a human verifier has reviewed the search log, including the search date, languages, sources, query terms, candidates inspected, closest match and any separating core rule. A search result is therefore evidence for a D classification only after human verification.

3.9 Eligibility

A game is core-eligible for a stated target class when it passes the gates, has MC ≥ 1 and CT-a ≥ 1, and the human-adjudicated carrying mechanic survives under that target class. For adaptations with an explicit educational objective, LOM ≥ 1 is applied as an additional pedagogical-suitability filter. No total score is reported. A failure on MC or CT-a indicates that eligibility is not supported by the currently documented evidence. It should not be interpreted as evidence that the traditional game itself is inherently unsuitable for adaptation. Additional documented evidence may support later human re-adjudication. In the pilot, LOM was coded against a domain-general educational objective. Neither coder rated any game below 1, so the LOM minimum did not affect any pilot verdict.

4. The workbench: Instrument-guided extraction with human adjudication

Applying the instrument manually across 518 entries entails a long coding campaign, in which consistent application may be difficult to maintain. We built a browser-based human-in-the-loop workbench in which a language model proposes a structured draft record and a human coder adjudicates the substantive field values (Figure 1). The language model does not determine eligibility: eligibility and the hybrid flag are derived deterministically only from the human-adjudicated record. The model draft, human-adjudicated record and derived verdict are therefore treated as distinct layers. A fuller account of the workbench architecture, prompt and schema design, provenance logging, and human–human and model–human evaluation is provided in the companion pipeline paper (Taheri and Büber Kaya, unpublished report).

4.1 Extraction

For each game the model receives a single prompt containing the instrument text (gates, route tests, every level descriptor, the ledger vocabulary, the distinctiveness levels) together with the entry and a fixed JSON schema. Where the serving model supports grammar-constrained decoding the schema is enforced during generation rather than validated afterwards, which removes malformed records as a class. There is no memory across games, so every record is produced under identical conditions; the instrument version and model tier are stored with the record. All drafts in this study were produced by Qwen3.5-35B, served locally. The model returns the record components in a fixed order: carrying mechanic; gates with reasons; primary and, where applicable, secondary route; the four scored criterion fields and two descriptive CT fields; the S/M survival profile together with the separate L setup/feasibility field; a transformation ledger; a distinctiveness search hypothesis without a final D category; source ambiguities; and a self-reported confidence field naming the least certain judgements. Each judgement is paired with a quotation from the entry. Where the source is silent the field goes to the ambiguity list and confidence drops. The order is deliberate: the route is fixed before the survival profile because it decides which platform question applies.

4.2 Adjudication

The draft renders as an editable sheet. A coder reviews it section by section, changes any field, and marks each section *reviewed.* The model cannot set that flag. The unedited draft is stored beside the adjudicated record and every difference between them is listed on demand. Two decisions are reserved to humans by construction. Distinctiveness is retrieval-dependent: the source-only extraction stage produces only a search hypothesis, not a D classification. The workbench runs retrieval as a separate logged step; candidate counterparts are then reviewed by a human, who assigns the final D category and verified status. The fuller retrieval procedure and its evaluation are reported in the companion pipeline paper (Taheri and Büber Kaya, unpublished report). And no record contributes to any count until every section has been reviewed.

4.3 Derived fields and queries

Eligibility and the hybrid flag are computed by rule from the adjudicated record, without weights and without model input. The pool is queried by route, survival under a target class and review status; where an educational objective has been specified it can additionally be filtered by LOM, and verified distinctiveness categories can be used for contextual analysis. The result is a candidate set per platform class with its reasons visible; the pool is not ranked.

4.4 What is logged

Instrument version; model tier and prompt hash; raw model draft; human-entered field values; associated source quotations, rationales and ledger notes with origin metadata; per-field differences between model and human values;

coder initials and date; second-coder record where present; search log for D. These are the data for the two evaluations in Section 6.

5. Worked examples

5.1 Seksek (hopscotch family): A maximum score that fails on screen

Seksek scored 10/10 under the first rubric. Under the instrument its carrying mechanic is hopping on one foot through a numbered grid in sequence, skipping the square holding the stone, without touching a line. Table 3 shows the ledger for class S. The carrying row is *substituted*: a tap-through of squares keeps the grid and loses the hop. Under the current instrument that substitution goes to the test of Section 3, which it fails on the second condition. The hop is the competence the game exists to exercise, and tapping does not transfer it, so Seksek is not eligible as a screen worked example. Under class M, step detection transforms the carrying action and the record therefore flags Seksek as a hybrid candidate. A co-located real-grid configuration with the device acting as caller or referee may additionally be recorded as a feasible support setup, but L does not carry the survival gate or hybrid flag. The first rubric, which scored Seksek 10/10, could not have produced this result, and the reason can be traced row by row in Table 3.

5.2 Akdört (four split-stick dice): Draft and independent coding compared

Akdört is a two-player race to fifty points over three sets, scored by the pattern of pale and bark faces shown by four thrown split sticks.[25] We ran the entry through the pipeline. The draft assigned R1 with no secondary route, MC 2, LOM 1, AGE 1, CT-a 2, the carrying mechanic *kept* under the screen and co-located classes and *transformed* under motion, a five-row ledger, D1 with "Kös Oyunu/Kös-Kos variants" as the counterpart (a variant name the entry itself supplies, not an international counterpart), and three source ambiguities. Overall confidence 0.85, with G3 and CT-c named as its least certain fields.

The two records differ on two fields (Table 4), and the more consequential is the route. Akdört illustrates why primary and secondary routes are distinguished. The coder's reading is consistent with the revised ordered test, which assesses the making phase first: what the sticks are made of governs how they fall, and therefore the score; competitive structure can be recorded separately as a secondary route where applicable. A parallel working brief had assigned Akdört to a construction category by lexical signal; the coder's reading agrees with that brief, but reaches it by a stated test rather than by vocabulary. The second difference is S survival: the model recorded the throwing action as *kept* on a screen, the coder as *transformed*, physical skill giving way to device logic while the probability distribution carries across.

Table 3. Seksek, transformation ledger, class S.

Element	Type	Original	Under S	Operation
Numbered grid	Material	Chalk on ground, 6–10 squares	Drawn grid	Kept
Throw stone to square *n*	Rule	Underarm, must land inside	Flick to square	Transformed
Hop through grid on one foot	Rule (carrying)	Balance and sequence; line contact ends the turn	Tap squares in order	Substituted
Skip the square holding the stone	Rule	Must jump over it	Square disabled	Kept
Retrieve stone on return	Rule	Bend while balancing	Auto-collect	Removed
Line judge	Social	Other players call faults	System detects	Substituted
Progress marker	Material	None	Level counter	Added

Table 4. Akdört: where the model's ratings and the coder's independent ratings differ. The coder rated the entry without consulting the draft, so these are differences between two readings, not edits applied to one.

Field	Model	Coder
route.primary	R1 contest	R3 construction
dig. S.op	Kept	Transformed

Everything else in the record is the model's, unchanged, and this is the main lesson of this example. The coder's ratings are hers; the reasons, quotations and ledger notes beside them are the draft's, because the coding sheet had a column for her value and a column for a free-text note but no column for her reason. Where she wrote a reason it went into the note, which is not a field the record has. For G1 her quotation and the model's are different passages of the same entry, and the record carries the model's. A record described as the coder's is therefore a record of her *ratings* carrying the model's supporting prose. This is a property of the sheet, not of the coding, and one that any study using this pipeline has to design out.

5.3 Nine bodily games: What the instrument rejects

A second working list of nine games from the corpus (chases, mounted games, wrestling, leapfrog, stick games, a gravity-car race) was chosen for distinctiveness rather than screen-portability, and every one would have been vetoed by the first screening's keyword pass. Run through the instrument in the formative pilot, these nine cases were subsequently coded independently by a second coder; the resulting agreement figures are reported in the companion study (Taheri and Büber Kaya, unpublished report) and are used here to identify fields requiring descriptor clarification rather than to estimate the reliability of the final instrument. Three fail gate G2: two because the rules require horses (Atlı Tura, Kökbörü), one because they need a field of over a hundred metres (Deynek Tenme); three receive the lowest developmental-adaptability category under the pilot descriptor, their competence tied to a body size or strength a younger player does not have (Araya Aldırmaca, Aşırtmalı Aba Güreşi, Atlı Tura); three lose the carrying mechanic on a touch screen but transform it under sensors, and are flagged as hybrid candidates (Aşırtmalı Aba Güreşi, Atlı Tura, Kökbörü); and one, Birdir Bir, loses it under both and fails gate G3 outright. **Four remain eligible on screen.** The Birdir Bir verdict and the hybrid flags follow the re-checked motion values reported in the companion paper (Taheri and Büber Kaya, unpublished report); under the instrument's first version Birdir Bir was itself a hybrid candidate, and the criterion that rejects it was tightened after the pilot found it too permissive. Route assignment by test disagreed with the list's own keyword categories on four of nine, each time because the test read what the rules require rather than the entry's vocabulary. Decisions of this kind are required of any instrument that can reject, and the first rubric could make none of them (Section 2).

A separate transformation analysis of the same games.

Eight of these nine games were also the subject of a working document dated 15 September 2026, written under a mobile-adaptation framing rather than the instrument's. The dating rests on the email record: the analysis was sent on 15 September and the first model drafts on 17 September, so it predates them. For each game it records what is *preserved*, what is *transformed* (as element-to-element mappings), what is *lost*, and what the digital version *adds*, together with the caution that "none of these added systems should be represented as authentic rules of the traditional game". Those four buckets are the ledger vocabulary of Section 3, arrived at independently, which is some evidence that the vocabulary is one a designer reaches for rather than one the instrument imposes. Table 5 reproduces it beside the model's later S survival value.

The two agree on seven of the eight. The exception is Birdir Bir, where the model drafted *kept* while the analysis preserves the roles, turn order, chants and round progression but records as lost "true physical proprioception and muscular elasticity, tactile weight-bearing contact with peers", which, if the carrying mechanic is the vault itself, implies *substituted.* That is a disagreement about what the carrying mechanic of a leapfrog game *is*, which is the kind of question the adjudication step exists to settle.

This is a weaker form of corroboration than it may appear. The document is by the first author, who is also a coder, so it shows one expert's judgement holding across two framings and two weeks, not agreement between two people. It is therefore not reliability evidence and is not treated as such.

6. Evaluation protocol

Two quantities must be reported separately, because they answer different questions.

6.1 Rubric reliability

Inter-rater reliability and validity address different questions.[26] Inter-rater reliability concerns the consistency with which independent coders apply the instrument and can be examined through field-level agreement statistics such as percentage agreement and Cohen's κ.[18,27] It does not by itself establish the validity of the intended selection inference. The nine-game pilot is therefore used formatively to identify descriptor weaknesses. The threshold for descriptor revision ($\kappa < .60$) was fixed in the study's coding protocol before the second coder began work. After those descriptors are revised and instrument v1.0 is frozen, inter-rater reliability will be evaluated on a new independent sample of 75 records drawn from the 518 entries under a sampling specification that uses no output of the instrument being evaluated. Neither gate results,

Table 5. The first author's transformation analysis of eight of the nine games. It was written on 15 September 2026 under a mobile-adaptation framing and before any model draft existed. The four buckets are her own; they are the instrument's ledger vocabulary, reached for independently. The final column is what the model later drafted for S survival.

Game	Preserved	Transformed (first row of each)	Lost	S
Araya Aldırmaca	Two-team structure, asymmetric runner/defender roles, central target …	Physical running → Virtual joystick movement …	Anaerobic sprinting, raw physical agility, somatic balance, …	Substituted
Atlı Tura	Two-team structure, explicit target selection, raid into enemy zone, …	Physical horse handling → Touchscreen steering …	True equestrian horsemanship, rider-horse biofeedback, …	Substituted
Bostan Oyunu	Two-team structure, twin garden sectors, simultaneous action, attacking …	Physical dirt plot → 2D/2.5D arena asset …	Sensory tactile feel of damp mud, hand-molding wet soil, …	Substituted
Kökbörü	Mounted locomotion, two-team competition, central ulak, ground pickup, …	Physical riding → Touchscreen vector …	True animal-rider biofeedback, physical muscle strain, …	Substituted
Birdir Bir	Ebe and vaulter roles, sequential turn taking, numbered round …	Physical running → Automated sprint/ timed …	Raw cardiovascular exertion, true physical proprioception and …	Kept
Deynek Tenme	Personal batons, sequential casting, distance ranking, precision …	Physical arm swing → Touchscreen drag-flick …	Biomechanical muscle fatigue and arm whip, visceral haptic …	Substituted
Akdört Oyunu	Four wooden staves, binary white/bark distinction, willow/poplar …	Physical branch → High-resolution 3D …	Sensory tactile grain of raw green wood, aromatic smell of …	Kept
Tahta Araba Yarışı (Formulaz)	Downhill mountain slope, pure gravity propulsion, foot-braking …	Physical foot friction → Touchscreen pressure …	Bodily centrifugal forces, biological fear of high-speed …	Substituted

criterion scores, route labels, survival operations, eligibility, nor the model drafts may inform selection; the sample is built from descriptive proxies read from the entry text and from semantic coverage of the corpus, with rare and boundary cases deliberately included and a seeded random component as a safeguard. The 75 remain balanced across the three adaptation routes (25 each), but the strata are route *proxies* taken from the entry — a making phase, an opponent or team contest, otherwise solitary play — not coded or drafted routes. Routes are assigned later by the coders under the frozen instrument, and the agreement between proxy stratum and coded route is reported as a diagnostic. If the closure-proxy stratum holds fewer than 25 entries, all are included and the remainder is split between the other two, with the actual split reported. Because selection uses no instrument output, the sample can be frozen before the model drafts it. The nine pilot games, the anchor games named in the level descriptors, and the worked example are excluded, since both coders have already seen them. An intentionally variance-enriched sample of this kind supports statements about application consistency, not about how common any outcome is in the corpus. In the 75-game study, reliability will be reported field by field using raw agreement together with an agreement coefficient appropriate to the measurement level and its uncertainty interval.[18,27] The number of applicable paired ratings will be reported for each field, and undecidable or missing ratings will be excluded only from that field's denominator rather than assigned a value. These analyses are intended to characterise application consistency under the frozen instrument, not the validity of the selection inference. Evidence supporting the validity of the intended selection use must be argued separately from reliability.

Agreement on the carrying mechanic will also be evaluated because this formulation conditions the subsequent survival profile and, potentially, the eligibility verdict. Each coder will independently state the carrying mechanic and the game-defining competence or relation embodied in it. Because these are free-text formulations rather than categorical ratings, they will not be forced into a Cohen's κ analysis. Instead, the two formulations will be compared under a pre-specified equivalence rule: equivalent when they identify the same game-defining relation or competence despite wording differences; partially overlapping when they share that core but one adds or omits a consequential component; and non-equivalent when they identify different game-defining relations or competences. We will report the proportion in

each category and whether non-equivalent formulations lead to different S/M survival classifications or eligibility verdicts.

6.2 Extraction accuracy

In the pilot, the human ratings were completed with the model values hidden and were compared with the model outputs only after coding was complete. Model–human agreement is computed per applicable field. For categorical criterion values, gate outcomes, route and S/M survival operations, the comparison is an exact match. For free-text fields it is reviewer-based, and for the ledger it is row-level. Source-only drafts are not evaluated on a final D category because distinctiveness requires retrieval and human verification. This comparison evaluates the assistive extraction system, not the inter-rater reliability or validity of the selection instrument. The fuller analysis is reported in the companion pipeline study (Taheri and Büber Kaya, unpublished report).

6.3 Corpus results

The pool is described by route distribution, gate rejection rate, survival profile per class, LOM distribution, and distinctiveness levels. The number of games rejected, with reasons, is the test of the claim that the instrument can reject. The corpus description is not reported here. The fifty shortlisted games were drawn under a literacy objective with a different rubric, so rescoring them would report a rejection rate belonging to that earlier filter rather than a test of this instrument. A corpus result has to come from a pool drawn under the instrument's own question, which is the stratified 75-game sample described in the companion paper (Taheri and Büber Kaya, unpublished report). What the pilot already supplies is the claim this paper makes: the first rubric could reject nothing, and the instrument declines five of nine: three at gate G2, one at G3, and four on the screen-survival rule, with several games failing on more than one ground.

7. Discussion

7.1 What the model is for

The workbench presents each entry to the model under the same instrument, schema and logging conditions. This standardises the extraction procedure for source quotation, ledger drafting and ambiguity recording. This procedural standardisation should not, however, be interpreted as evidence that the model's substantive judgements are themselves consistent. The instrument exists to make these judgements explicit: whether a learning behaviour is intrinsic, whether a mechanic survives a class, and whether a counterpart exists. The first question to ask of a rubric is whether two humans agree when applying it. That is reliability evidence, which does not by itself establish validity. Using the model as a second coder would measure the model, not the rubric. We therefore keep the two evaluations apart.

7.2 Where the model is most likely wrong

The field most at risk is distinctiveness. Asked without retrieval, the model returned no judgement on this field at all: it recorded D1 for every game in the pilot, which is the value printed in the output skeleton it was given, and twice recorded D1 with an empty separatingRule, which that level's definition forbids. The field is a hypothesis by construction, yet in the pilot the model did not supply even that. The workbench now answers it in a separate step: a web search supplies the candidate sources and the model selects among them by index, so every URL in the record came from the search engine rather than from the model. The level and the *verified* flag remain the coder's. This removes direct model fabrication of source URLs, but it does not remove retrieval, relevance or classification error; the final D category therefore remains a human-verified decision. The fuller retrieval procedure and its evaluation are reported in the companion pipeline study (Taheri and Büber Kaya, unpublished report).

7.3 Screen versus hybrid

The survival profile turns "can it be digitised", which is true of everything, into "what survives on which interface". In our pool the games with the richest bodily content (chases, mounted games, leapfrog) are lost under S and transformed under M, which describes their fit to a platform and says nothing about their quality. The instrument records that rather than penalising it, and the hybrid flag is how a screen-first project can still say something defensible about them.[8,10]

7.4 Relation to Ludii

Ludii[4] formalises rules for board, card, dice and tile games and excludes physical skill. The ledger vocabulary is a transformation record over rule and material elements including bodily and social ones. The two are complementary, and games that are kept under S are the natural candidates for a Ludii description.

7.5 What we do not claim

Earlier work already covers several adjacent forms of digital and hybrid adaptation. The Head Up Games programme combines traditional and digital play[8]; interactive tag systems augment physical chasing games with sensing or projected layers[28]; heritage-oriented frameworks have supported expert-configurable exergame creation[29]; interactive playgrounds

have been designed from traditional children's play[30]; sensor- and mobile-based systems have connected outdoor play with digital games[31]; mixed-reality playgrounds have combined physical play with robotic interaction[32]; and VR methods have been proposed for reconstructing a traditional sport as heritage.[33] In Türkiye, traditional children's games have been carried into digital contexts through a pandemic-era digitally mediated transfer project[34] and a national digital atlas.[35] Related distinctions between a game's retained core, its adaptation to a new environment and its physical or digital version already exist,[15,36,37] and the five-operation ledger vocabulary used here is a synthesis drawing on.[15,24]

What is claimed is narrower: an instrument applied to a documented game *before* any adaptation exists, whose output is a record rather than a score, and which can return a reasoned rejection. The claim therefore rests on rejection rather than novelty: the instrument rejects candidates and, when applied, it exposed defects in some of its own criteria.

8. Limitations

The corpus is one national encyclopaedia. The entries vary in the detail of their rules, and the ambiguity list partly measures the source rather than the game. The pilot's first coding was by a single coder. A second coder has since coded the same nine games from the entries alone, and the resulting κ values, below the revision threshold on all four scored criteria (MC, LOM, AGE and CT-a) and on the route field, are reported in the companion paper; at nine games the intervals are wide, so those figures locate weak descriptors rather than measure reliability. Extraction accuracy on the nine records is reported in the companion paper (Taheri and Büber Kaya, unpublished report) and is indicative only at $n = 9$. Each record keeps the independently produced human field values apart from the model draft, with source quotations and supporting notes marked by origin (see Data availability). Community relation (CT-c) and provenance obligations are recorded, not evaluated, and the ethics of adapting some bodily games (contact wrestling, whip strikes) for a classroom setting are outside the instrument. Further limits concern the scope of the decision the instrument produces. Eligibility is a pre-adaptation judgement. It is not evidence that a resulting digital or hybrid game will be pedagogically effective, engaging, usable, safe, affordable, or acceptable to the community from which the game is documented; those claims require participatory design, prototyping and evaluation with intended players and relevant knowledge holders.

Several judgements remain interpretive. Identifying a game's carrying mechanic is consequential: defining a bodily action, a strategic problem or a social relation as the carrying one can change the recorded survival profile and therefore the verdict. Because the carrying-mechanic formulation conditions all subsequent survival judgements, the 75-game reliability study will evaluate agreement on this field in addition to the scored and categorical fields. Agreement on the free-text formulations will be assessed under the pre-specified equivalence procedure described in Section 6, and disagreements will additionally be examined for whether they propagate to different survival or eligibility decisions.

The interface classes are abstractions, not deployment specifications. Sensor accuracy, latency, camera placement, access to suitable devices and space, cost, accessibility, safeguarding and the practical limits of co-located sensing are not represented in them. Class L has been removed from the gate because it could not discriminate: under a co-located setup the physical game is unchanged by definition, so L cannot show that a game survives digitisation. It is retained as a hybrid design option and described in the companion framework paper.

9. Conclusion

A selection instrument for traditional games should be able to return a reasoned non-selection decision, state under which interface class the carrying mechanic survives, and keep distinctiveness separate from suitability. The instrument described here does those things at the level of the formative cases: it produces traceable, non-compensatory decisions that the earlier rubric could not, and its application exposed weaknesses in both a gate and several descriptors. These findings establish functional discrimination and auditability in the development set, not a corpus-level rejection rate and not the reliability of the frozen instrument. The agreement weaknesses identified in the nine-game pilot therefore serve as targets for revision before the independent 75-game reliability study.

Ethics and consent

This study analysed written descriptions of traditional games and involved no human participants, no personal data and no intervention. Ethical approval and consent were not required.

Software and code availability

- Software available from:
- https://github.com/Taheri-meisam/carrying-mechanic-workbench/releases/tag/v1.1.0
- Source code available from:

- https://github.com/Taheri-meisam/carrying-mechanic-workbench
- Archived source code at time of publication:
- https://doi.org/10.5281/zenodo.23005253[39]
- Taheri M, Büber Kaya H. Carrying Mechanic workbench: a selection instrument and workbench for traditional games (Version 1.1.0) [Software]. Zenodo; 2026. doi:10.5281/zenodo.23005253
- **License:** MIT (OSI-approved)

The repository carries the instrument, the prompt, the schema and the analysis scripts, but not the coded records, which are available on request as described under Data availability; the values behind the figures reported in this article are in the underlying data.[38] The test suite runs on a synthetic corpus and makes no model calls.

Use of generative AI

Qwen3.5-35B, served locally, was used as part of the research system to generate the structured draft records evaluated in this study. These model-generated drafts were kept separate from the independently produced human ratings, as described in the Methods. Claude (Anthropic; models Claude Opus 5 and Claude Fable 5.1, used through Claude Code in September 2026) was used separately during manuscript preparation for drafting and language editing, and for programming assistance in developing the research software, in order to shorten the time that writing and implementation took; it was not the model evaluated in the research pipeline and did not supply the human-coded ratings reported in this article. The authors reviewed and revised all AI-assisted material and take responsibility for the content.

Data availability

Underlying data

Zenodo: Underlying data for "Choosing What Not to Adapt: An Instrument That Can Say No Before Traditional Games Are Digitised for Screen or Hybrid Play". https://doi.org/10.5281/zenodo.23280520[38]

This project contains the following underlying data:

- screening_funnel.csv: counts at each stage of the first screening (Section 2)
- first_rubric_scores_50.csv: first-rubric scores of the 50-game pool (Section 2, Table 1, Section 5.1)
- akdort_model_vs_coder.csv: model draft and coder's ratings for Akdört (Section 5.2, Table 4)
- pilot_nine_games_adjudicated.csv: instrument codes for the nine pilot games (Section 5.3)
- pilot_nine_games_two_coders.csv: the two coders' independent ratings of the nine pilot games, from which the agreement figures in the Abstract and Section 8 are computed
- synthetic_corpus_24.json: 24 game descriptions written by the authors, for running the software
- README.md: which file reproduces which figure, and definitions of all variables

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Restricted data

The coded records contain the full text of, and extended quotations from, the English translations described below, whose rights are held by the Ministry of Youth and Sports of the Republic of Türkiye. The Ministry's written authorisation (29 June 2026) covers research use but not redistribution, so the records cannot be deposited in an open repository. The authors obtained the translations from the Ministry, and a reader can obtain them by the same means: requests for access can be sent to the corresponding author and will be forwarded to the Ministry, which decides on access and on its conditions.

Source data

The game descriptions coded in this study are entries of the *Encyclopedia of Traditional Sports and Games*,[25] an online reference work whose Turkish entries have undergone editorial review for a forthcoming print edition. The coding used English translations of these entries prepared for a forthcoming English edition by the Ministry of Youth and Sports of the Republic of Türkiye (unpublished), supplied by the Ministry for this project in advance of publication. The authors neither generated nor own this material and cannot redistribute it. The quotations reproduced in this article are taken from these translations and used under the Ministry's written authorisation.

Publicly available representative data

The Turkish entries of the encyclopaedia are publicly available online at https://encyclopedia.worldethnosport.org. They are the entries from which the translations were made, so the instrument can be applied to them as described in this article. A synthetic corpus of 24 game descriptions written by the authors, on which the pipeline and the analysis scripts can be run without the restricted text, is included in the underlying data (CC-BY 4.0) and in the software repository (MIT licence).

Acknowledgements

This publication is based upon work from COST Action CA22145 GameTable (Computational Techniques for Tabletop Games Heritage), supported by COST (European Cooperation in Science and Technology). The work was carried out under a Short-Term Scientific Mission of that Action. We thank Fatma Özlem Özkan of the Turkish Ministry of Youth and Sports, who granted access to the encyclopaedia and served as the independent second coder for the pilot.

References

1. Sallnäs Pysander EL, Mårtensson F, Waern A, *et al.*: **Nature and digitalization challenging the traditional playground.** *Urban For. Urban Green.* 2024; **93**: 128148.
Publisher Full Text

2. Büyükokutan Töret A, Baştürk N: **Mangalanın dijital kültür ortamındaki seyri Üzerine.** *Kültür Araştırmaları Dergisi.* 2023; **19**: 41–59.
Publisher Full Text

3. Browne C: **A class grammar for general games.** *Computers and games (CG 2016).* Springer; 2016; p. 167–82. (Lecture notes in computer science).
Publisher Full Text

4. Piette É, Soemers DJNJ, Stephenson M, *et al.*: **Ludii – the ludemic general game system.** *ECAI 2020.* IOS Press; 2020; p. 411–8. (Frontiers in artificial intelligence and applications).
Publisher Full Text

5. Crist W, Stephenson M, Piette É, *et al.*: **The ludii games database: A resource for computational and cultural research on traditional board games.** *Digital Humanities Quarterly.* 2024; **18**(4).
Publisher Full Text

6. Mueller F, Edge D, Vetere F, *et al.*: **Designing sports: A framework for exertion games.** *Proceedings of the SIGCHI conference on human factors in computing systems.* 2011; p. 2651–60.
Publisher Full Text

7. Mueller FF, Gibbs MR, Vetere F, *et al.*: **Designing for bodily interplay in social exertion games.** *ACM Transactions on Computer-Human Interaction.* 2017; **24**(3): 1–41.
Publisher Full Text

8. Soute I, Markopoulos P, Magielse R: **Head up games: Combining the best of both worlds by merging traditional and digital play.** *Pers. Ubiquit. Comput.* 2010; **14**(5): 435–444.
Publisher Full Text

9. Söbke H, Baalsrud Hauge J, Stefan IA: **Prime example Ingress: Reframing the pervasive game design framework (PGDF).** *Int. J. Serious Games.* 2017; **4**(2).
Publisher Full Text

10. Tisserand Y, Magnenat-Thalmann N, Unzueta L, *et al.*: **Preservation and gamification of traditional sports.** *Mixed reality and gamification for cultural heritage.* Springer; 2017; p. 421–46.
Publisher Full Text

11. Gunter GA, Kenny RF, Vick EH: **Taking educational games seriously: Using the RETAIN model to design endogenous fantasy into standalone educational games.** *Educ. Technol. Res. Dev.* 2008; **56**(5-6): 511–537.
Publisher Full Text

12. Arnab S, Lim T, Carvalho MB, *et al.*: **Mapping learning and game mechanics for serious games analysis.** *Br. J. Educ. Technol.* 2015; **46**(2): 391–411.
Publisher Full Text

13. Petri G, Gresse von Wangenheim C, Borgatto AF: **MEEGA+, systematic model to evaluate educational games.** *Encyclopedia of computer graphics and games.* Springer; 2018.
Publisher Full Text

14. Fu FL, Su RC, Yu SC: **EGameFlow: A scale to measure learners' enjoyment of e-learning games.** *Comput. Educ.* 2009; **52**(1): 101–112.
Publisher Full Text

15. Mesquita CB, Jucá PM: **G.A.M.E.: Guide for the adaptation of modern board games to digital environments.** *Journal on Interactive Systems.* 2026; **17**(1): 531–546.
Publisher Full Text

16. Pasch M, Bianchi-Berthouze N, van Dijk B, *et al.*: **Movement-based sports video games: Investigating motivation and gaming experience.** *Entertainment Computing.* 2009; **1**(2): 49–61.
Publisher Full Text

17. Brookhart SM: **Appropriate criteria: Key to effective rubrics.** *Frontiers in Education.* 2018; **3**: 3.
Publisher Full Text

18. Jönsson A, Svingby G: **The use of scoring rubrics: Reliability, validity and educational consequences.** *Educ. Res. Rev.* 2007; **2**(2): 130–144.
Publisher Full Text

19. Habgood MPJ, Ainsworth SE: **Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games.** *J. Learn. Sci.* 2011; **20**(2): 169–206.
Publisher Full Text

20. Hirsh-Pasek K, Zosh JM, Golinkoff RM, *et al.*: **Putting education in 'educational' apps.** *Psychol. Sci. Public Interest.* 2015; **16**(1): 3–34.
PubMed Abstract | Publisher Full Text

21. Mortara M, Catalano CE, Bellotti F, *et al.*: **Learning cultural heritage by serious games.** *J. Cult. Herit.* 2014; **15**(3): 318–325.
Publisher Full Text

22. UNESCO: **Convention for the safeguarding of the intangible cultural heritage.**
Reference Source 2003.

23. Sousa M: **Finding ludemes in modern board games: Analyzing the top number one games of board game geek.** *Board Game Studies Journal.* 2023; **17**(1): 205–229.
Publisher Full Text

24. Sousa M, Zagalo N, Oliveira AP: **Mechanics or mechanisms: Defining differences in analog games to support game design.** *2021 IEEE conference on games (CoG).* 2021; p. 1–8.
Publisher Full Text

25. World Ethnosport Confederation: **Encyclopedia of traditional sports and games [geleneksel sporlar ve oyunlar ansiklopedisi] Online reference work.**
Reference Source

26. Messick S: **Validity.** Linn RL, editor. *Educational measurement.* New York: American Council on Education; Macmillan;3rd edn. 1989; p. 13–103.

27. Hallgren KA: **Computing inter-rater reliability for observational data: An overview and tutorial.** *Tutor. Quant. Methods Psychol.* 2012; **8**(1): 23–34.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Moreno A, Poppe R: **Automatic behavior analysis in tag games: From traditional spaces to interactive playgrounds.** *Journal on Multimodal User Interfaces.* 2016; **10**(1): 63–75.
Publisher Full Text

29. Grammatikopoulou A, Laraba S, Sahbenderoglu O, *et al.*: **An adaptive framework for the creation of exergames for intangible cultural heritage (ICH) education.** *J. Comput. Educ.* 2019; **6**(3): 417–450.
Publisher Full Text

30. Tetteroo D, Reidsma D, van Dijk B, *et al.*: **Design of an interactive playground based on traditional children's play.** *Intelligent technologies for interactive entertainment (INTETAIN 2011).* Springer; 2012; p. 129–38. (LNICST).
Publisher Full Text

31. Díaz DJ, Boj C, Portalés C: **HybridPLAY: A new technology to foster outdoors physical activity, verbal communication and teamwork.** *Sensors.* 2016; **16**(4): 586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

32. Lupetti ML, Piumatti G, Germak C, *et al.*: **Design and evaluation of a mixed-reality playground for child-robot games.** *Multimodal Technologies and Interaction.* 2018; **2**(4): 69.
Publisher Full Text

33. Gaugne R, Barreau JB, Duc-Martin P, *et al.*: **Sport heritage in VR: Real tennis case study.** *Frontiers in Virtual Reality.* 2022; **3**: 922415.
Publisher Full Text

34. Özdemir M: **Covid-19 salgın sürecinde geleneksel oyunların dijitalde aktarımı: Evde oyun var!.** *Folklor/Edebiyat.* 2022; **28**(111): 661–680.
Publisher Full Text

35. İneç ZF: **Kültürü sayısal bir ağa dönüştürmek: Türkiye geleneksel çocuk oyunları dijital atlası örneği.** *Iğdır Üniversitesi Sosyal Bilimler Dergisi.* 2022; **30**: 252–269.
Publisher Full Text

36. Grabarczyk P, Aarseth E: **Port or conversion? An ontological framework for classifying game versions.** *DiGRA Digital Library.* 2019.
Publisher Full Text

37. Rogerson MJ, Sparrow LA, Gibbs MR: **Unpacking 'boardgames with apps': The hybrid digital boardgame model.** *Proceedings of the 2021 CHI conference on human factors in computing systems.* 2021.
Publisher Full Text

38. Büber Kaya H, Taheri M: Underlying data for 'Choosing What Not to Adapt: An Instrument That Can Say No Before Traditional Games Are Digitised for Screen or Hybrid Play'. [Dataset]. *Zenodo.* 2026.
Publisher Full Text

39. Taheri M, Büber Kaya H: **Carrying Mechanic workbench: A selection instrument and workbench for traditional games [Software].** *Zenodo.* 2026.
Publisher Full Text