

Peer-Review

Esboña, Alonso J. 2026. "Designing Believable Worlds: NPC AI and Immersive Evolution in The Legend of Zelda Series." *Journal of High School Science* 10 (3): 414–34. <https://doi.org/10.64336/001c.166485>.

I enjoyed reading the manuscript. You have actually conceptualized a framework with which to compare the Zelda-progression with each game. Problem is, you have not quantified your 7 dimensions, hence the paper is not evidentiary. Your idea is better than your methodology. Here's what I suggest that will make this a very interesting contribution. Figure out a way to measure those dimensions.

I am attaching a pdf file that contains details on how to go about this. The content is self-explanatory. If you can operationalize those dimensions, I believe you have the makings of a good manuscript.

Also make sure that you have a defensible sampling protocol and that you sample at least 3 times (say in the beginning of the game, the middle and then at the end of the game).

I look forward to reviewing your revised paper.

Response to Reviewers

Manuscript: Designing Believable Worlds: NPC AI and Immersive Evolution in The Legend of Zelda Series

Author: Alonso J. Esboña (student); Rebecca Larkin (mentor)

Journal: Journal of High School Science

[BEFORE SUBMITTING] Add the Scholastica manuscript ID and the resubmission date on this line.

Note: all page numbers below refer to the revised manuscript.

Dear Dr. Apte,

Thank you for the opportunity to revise this manuscript, and thank you to both reviewers for their detailed comments. The reviews identified a fundamental weakness in the original submission: the seven analytical dimensions were presented as descriptive categories rather than measured variables. The revision addresses this directly. All seven dimensions have been operationalized as observable surrogate measures, observational data were collected at three sampling points in each title, and the results are now reported quantitatively and aggregated into an Environmental Coherence Index.

Each reviewer comment is reproduced verbatim below, followed by a description of how and where it has been addressed.

Reviewer 1

Comment 1 — full comment as submitted

"I enjoyed reading the manuscript. You have actually conceptualized a framework with which to compare the Zelda-progression with each game. Problem is, you have not quantified your 7 dimensions, hence the paper is not evidentiary. Your idea is better than your methodology. Here's what I suggest that will make this a very interesting contribution. Figure out a way to measure those dimensions. I am attaching a pdf file that contains details on how to go about this. The content is self-explanatory. If you can operationalize those dimensions, I believe you have the makings of a good manuscript. Also make sure that you have a defensible sampling protocol and that you sample at least 3 times (say in the beginning of the game, the middle and then at the end of the game). I look forward to reviewing your revised paper."

This comment contains three distinct requests, addressed separately below.

1(a) Quantify and operationalize the seven dimensions

Response:

The seven dimensions have been operationalized as observable behavioral surrogate measures, following the approach set out in the attached PDF. Because the internal AI source code of the three

titles has not been publicly documented, each dimension is measured through a countable manifestation of the underlying system rather than through the architecture itself.

Where addressed:

- A new Methods subsection, “Operationalizing the seven dimensions,” explains the surrogate-measure approach and why it was necessary. (p. 3)
- New Table 1 lists each dimension, its a priori ceiling value, and the exact surrogate measure that was counted. (p. 3)
- The Results section now reports measured raw and normalized values for every dimension rather than qualitative description. (pp. 6–12)
- New Table 2 reports raw and normalized scores for all seven dimensions across the three titles, together with the composite index. (p. 8)
- New Figure 1 presents the normalized dimension scores for the three titles. (p. 9)
- New Figure 2 presents the Environmental Coherence Index by title. (p. 11)
- New Figure 3 presents the dimension profile of each title. (p. 12)
- A composite Environmental Coherence Index was constructed from the seven normalized dimensions: 2.89 for Ocarina of Time, 4.22 for The Wind Waker, and 8.19 for Breath of the Wild, on a 0–10 scale. (pp. 8, 11)

1(b) Defensible sampling protocol

Response:

A sampling protocol has been defined and stated explicitly so that the observations can be reproduced by another researcher.

Where addressed:

- A new Methods subsection, “Sampling protocol,” specifies a fixed 30-minute observation window at each sampling point. (p. 5)
- Observations were restricted to a consistent NPC type across all three titles (ordinary town and village NPCs), so that the comparison is like-for-like. (p. 5)
- Player–NPC interaction was measured with a fixed set of five player actions applied identically in each title: approach, draw weapon, initiate dialogue, jump nearby, and remain idle. (p. 5)
- Believability was rated on a defined 1–5 appropriateness scale. (pp. 3, 5)
- A new Methods subsection, “Data collection,” states that measurements were taken from publicly available, continuous non-commentary gameplay recordings, which provide a timestamped and re-observable record that another researcher can verify against the same source. The three recordings are cited as references (13), (14) and (15). (p. 5)

1(c) Sample at least three times per game (beginning, middle, end)

Response:

Each title was sampled at three points: early game (approximately the first 10% of the main story), mid game (approximately 50%), and late game (approximately 90%). Every dimension was measured at each of the three points, and the reported dimension value is the mean of the three sampling points.

Where addressed:

- The “Sampling protocol” subsection defines the three sampling points. (p. 5)
- The “Normalization and the Environmental Coherence Index” subsection explains that dimension values are averaged across the three sampling points before normalization. (p. 5)
- Limitations notes that three sampling points per title were assumed adequate for a comparative rather than exhaustive analysis. (p. 16)

Reviewer 2

Comment 1

"Please verify that all links to the references point to the correct source."

Response:

Every reference has been checked and corrected.

- The reference list has been renumbered so that each in-text citation number corresponds to the correct entry, and every in-text citation was verified individually against the list. (p. 17)

- References are formatted consistently in APA style, with DOI links added where a DOI exists (Jennett et al.; Warpefelt and Verhagen; Feng and Tan). **(p. 17)**
- The three gameplay recordings used as data sources were added as references (13), (14) and (15), each with channel, video title, date and URL. **(p. 17)**
- Book and chapter references were corrected to name the publisher and editors rather than the hosting website. **(p. 17)**

Comment 2

"The authors must disclose and acknowledge any assistance received in the preparation of this manuscript, including but not limited to editorial, technical, analytical, or writing support. All such contributions must be clearly stated in the Acknowledgments section."

Response:

A disclosure statement has been added to the Acknowledgements section. It states the use of an AI assistant for editorial support, including feedback on structure, language and grammar editing, spreadsheet and formula construction for organizing the data, and generation of the figures from author-collected data. The statement also confirms that all gameplay observations and data were collected by the author, who reviewed all content and takes full responsibility for it. Mentor support is acknowledged separately.

- Disclosure statement added to the Acknowledgements section. **(p. 17)**
- A conflict of interest and funding statement was also added, declaring no funding and no conflicts of interest. **(p. 17)**

Comment 3

"Include enough recent references along with foundational ones. Ensure references directly support your claims. Avoid "padding." Use credible sources (peer-reviewed journals, reputable books, official reports)."

Response:

The reference list has been reviewed for relevance, quality and balance, and a primary developer source has been added.

- A primary developer source was added: the Game Developers Conference 2017 presentation by Fujibayashi, Takizawa and Dohta, the director, art director and technical director of Breath of the Wild. This is cited in support of the systemic and environmental behavior described in the Breath of the Wild case study and in the discussion of the results, replacing what were previously inferred claims with a developer-documented account. **(reference (16); cited pp. 7, 12)**
- Foundational technical sources are retained and cited at the specific claims they support: Millington on navigation meshes and decision-making architectures, Orkin on goal-oriented action planning, and Al Shehabi on finite state machines. **(pp. 2–3)**
- Theoretical sources are cited where their definitions are used: Jennett et al. for immersion and Warpefelt and Verhagen for believability. **(p. 3)**
- Non-peer-reviewed sources are now used only for reception and historical context and not to support technical claims about AI implementation. The IGN review is presented explicitly as critical reception. **(p. 7)**
- Claims that were previously asserted without support are now attached either to a cited source or to the author's own footage observations, with the source recording cited. **(pp. 6–12)**

Comment 4

"All assumptions (including implicit assumptions) must be explicitly stated and clearly justified. The authors should explain why each assumption is reasonable and discuss its impact on the results and conclusions."

Response:

Five assumptions are now stated explicitly and justified, and their impact on the conclusions is discussed.

- (1) Equal weighting of the seven dimensions, on the basis that no evidence supports differential weighting; equal weighting is the most transparent and conservative choice. **(p. 5)**

- (2) Ceilings defined a priori from the measurement instrument rather than from the observed data, so that the scale remains independent of this particular three-title sample. All seven ceiling values are given in Table 1. (pp. 3, 5)
- (3) Recorded footage assumed representative of typical NPC behavior at each sampling point, with the fixed 30-minute window used to reduce selective observation. (p. 5)
- (4) The Environmental Coherence Index measures observable environmental coherence and not immersion itself; immersion would require separate measurement through player studies. (p. 5)
- (5) Three sampling points per title assumed adequate for a comparative rather than exhaustive analysis. (p. 5)

In addition, a sensitivity check was carried out to test whether the conclusions depend on these assumptions. The index was recomputed using simple raw averaging of the unnormalized values (1.19, 1.44, 3.42) and using min–max normalization across the three games (0.05, 0.22, 1.00). The ranking of the three titles was unchanged under all three schemes, which indicates that the conclusion does not depend on the specific ceiling choices.

- Sensitivity check reported in the “Normalization and the Environmental Coherence Index” subsection. (p. 5)

Comment 5

"Avoid overreaching conclusions that extend beyond what is supported by the data and analysis."

Response:

Claim language has been revised throughout so that the stated conclusions match what the study actually measured. Specific changes include:

- The Abstract now reports the measured index values and states directly that the observed increase is consistent with improved immersion but that immersion itself was not directly measured. (p. 1)
- Evaluative phrasing was removed, including “masterclass,” “so close to real life,” “greatest of all time” and “demonstrates flawlessly.” (pp. 6–12)
- “Positive linear correlation” was replaced with “consistent increase,” since three data points cannot establish a correlation. (p. 12)
- Absolute language was softened, for example from “navmeshes are a must” to “well-suited,” and from “impossible with a scripted design” to “difficult to achieve.” (pp. 12, 16)
- A claim about the wider industry was narrowed to the three titles studied. (p. 12)
- A new subsection, “Alternative explanations,” was added to the Discussion. It acknowledges that art direction, level design and hardware could also contribute to perceived immersion, and explains that the seven measures were deliberately defined as behavior-specific events, so that the observed differences track NPC behavior rather than aesthetics. (p. 16)
- The Limitations section was expanded to state that data came from recorded footage rather than live play, that behavior outside the sampled windows was not captured, and that believability and emergence were rated by a single observer without inter-rater validation. (p. 16)

Comment 6

"Verify that you have used past perfect tense and third person throughout the manuscript wherever applicable."

Response:

The manuscript was revised for tense and person. All first-person pronouns were removed from the body of the paper and replaced with third-person constructions, and the study’s procedures and observations are reported in the past tense. First person is retained only in the Acknowledgements, where it is conventional.

- Verified across the Abstract, Introduction, Methods, Results, Discussion, Limitations and Conclusion. (pp. 1–16)

Summary of principal changes

Section	Change	Location
Abstract	Rewritten to report the measured index values and to state that immersion itself was not directly measured.	p. 1

Section	Change	Location
Methods	Four new subsections: operationalization of the seven dimensions, sampling protocol, data collection, and normalization with five stated assumptions and a sensitivity check.	pp. 3–5
Table 1	New. Each dimension with its a priori ceiling and its observable surrogate measure.	p. 3
Results	Rewritten to report measured raw and normalized values for each of the seven dimensions at three sampling points per title.	pp. 6–12
Table 2	New. Raw and normalized dimension scores and the Environmental Coherence Index for each title.	p. 8
Figures 1–3	New. Normalized dimension scores, Environmental Coherence Index, and dimension profiles.	pp. 9, 11, 12
Discussion	Claims tied to specific measured dimensions; developer source added; evaluative language removed.	pp. 12–16
Alternative explanations	New subsection addressing art direction, level design and hardware as competing explanations.	p. 16
Limitations	Expanded to cover footage-based collection, single-observer rating, hardware differences and inferred architectures.	p. 16
Future research	New section proposing player studies, application of the rubric to other franchises, and extension to a fourth title.	p. 16
Acknowledgements	Disclosure of assistance added; conflict of interest and funding statement added.	p. 17
References	Renumbered and verified; APA formatting; DOIs added; three gameplay data sources and one primary developer source added (16 references total).	p. 17

The author thanks both reviewers again for comments that substantially improved the manuscript.
Sincerely,
Alonso J. Esboña

Thank you for addressing my comments. Accepted with one caveat that can be included during copyediting:

The manuscript should explain how the a priori ceiling values (5, 6, 7, etc.) were derived during construction of the measurement instrument. Although the authors state that these values were specified before data collection, a brief justification for each ceiling would improve transparency and reassure readers that the normalization was not arbitrary.