Immersive Literary Learning Through Augmented Reality: Evaluating Cognitive Engagement Using a Rasch Measurement Approach

Onok Yayang Pamungkas[1,*], Achmad Fauzan[2], Lahan Adi Purwanto[2], Feisal Aziez[3] and Akhmad Fauzan[1]

[1]*Indonesian Language and Literature Education, Universitas Muhammadiyah Purwokerto, Indonesia*

[2]*Informatics Engineering, Universitas Muhammadiyah Purwokerto, Indonesia*

[3]*English Language and Education, Universitas Muhammadiyah Purwokerto, Indonesia*

Abstract: This study examines the measurement quality and cognitive engagement of students in augmented reality–based literary psychology learning using a Rasch model framework. A total of 52 undergraduate students participated in the implementation of an Augmented Reality Psychology Sastra (ARPS) application, which integrates visual-interactive features with literary psychological concepts. Data were collected through a 24-item dichotomous quiz designed to capture multiple levels of cognitive processing, including conceptual understanding, character interpretation, and inferential reasoning. Rasch analysis revealed acceptable measurement properties, with person reliability of 0.78 and item reliability of 0.91, indicating consistent responses and stable item calibration. Item fit statistics showed that all items functioned within acceptable limits (0.88–1.22), supporting construct validity. However, the Wright Map indicated a targeting mismatch, with mean person ability exceeding item difficulty (+0.42 logits), suggesting that the instrument was relatively easy for most participants. Person fit analysis further confirmed response consistency, with only 5.8% misfit cases. These findings suggest that augmented reality enhances cognitive performance through immersive and multimodal learning experiences, while also highlighting the need for more complex items to improve measurement sensitivity at higher ability levels. The study contributes to the integration of immersive learning technologies and psychometric validation in literary education contexts.

Keywords: Augmented reality learning, Literary psychology, Rasch model, Cognitive engagement, Measurement validity, Immersive education, Educational assessment.

1. INTRODUCTION

The transformation of 21st-century learning demands the integration of digital technologies that not only increase access, but also deepen the quality of learners' cognitive processing. In this context, augmented reality (AR) is emerging as a technology capable of delivering immersive learning experiences through the integration of visual and interactive representations, thereby enriching the construction of meaning and increasing learning engagement (Ibáñez & Delgado-Kloos, 2018; Radianti *et al*., 2020; Akçayır & Akçayır, 2017; Dunleavy & Dede, 2014). A number of studies have shown that AR is effective in facilitating conceptual understanding, especially in abstract and complex domains, through the presentation of multimodal information that reduces initial cognitive burden and improves retention (Bacca *et al*., 2014; Garzón *et al*., 2019, Garzón & Acevedo, 2021; Radu, 2014; Makransky & Petersen, 2019, Makransky *et al*, 2021). Within the framework of multimedia learning and dual coding theory, the combination of visual and verbal information allows for deeper and more structured cognitive processing (Mayer, 2009; Paivio, 1991; Sweller, 2011).

However, in the field of literary learning—especially literary psychology—the use of immersive technology is still relatively limited. Learning literary psychology requires a high level of interpretive skills, such as character analysis, inner conflict, and complex narrative symbolism (Eagleton, 2008; Wellek & Warren, 1949). The main challenge in this domain lies in its abstract and subjective nature, so it is often difficult for students to understand in depth (Bruner, 1990; Rosenblatt, 1994). Therefore, the integration of AR has the potential to be a pedagogical solution that is able to visualize psychological dynamics in a concrete way, thereby bridging the gap between literary texts and participants' cognitive understanding (Ibáñez & Delgado-Kloos, 2018; Garzón *et al*., 2019).

However, improving the quality of learning through immersive technology has not always been followed by the development of adequate evaluation instruments. Most studies still focus on improving learning outcomes, without paying attention to the validity and precision of the measurements used to evaluate these outcomes (Radianti *et al*., 2020; Garzón *et al*., 2019; Makransky *et al*, 2023). In the perspective of modern psychometrics, measurement validity is a construct that must be built through the suitability between data and models, not just reliability or raw scores (Messick, 1995; Kane, 2013). The Rasch model offers a robust

*Address correspondence to this author at the Indonesian Language and Literature Education, Universitas Muhammadiyah Purwokerto, Indonesia; E-mail: onokyayangpamungkas@gmail.com

Received on 01-03-2026 Accepted on 03-04-2026 Published on 14-04-2026

E-ISSN: 2978-8811/26

approach to ensure that the instrument is capable of producing invariant measurements, as well as providing diagnostic information regarding item quality and participant responses (Boone, 2020; Bond & Fox, 2015; Tennant & Conaghan, 2007; Linacre, 2002). However, studies that integrate Rasch-based evaluation in the context of AR learning, particularly in the domain of literary psychology, are still very limited.

Despite the rapid proliferation of immersive technologies in education, several critical gaps remain insufficiently addressed. First, existing studies on augmented reality and immersive learning environments predominantly emphasize learning outcomes and user engagement, while overlooking the robustness of measurement instruments used to capture these constructs. As a result, claims regarding the effectiveness of immersive learning often rely on raw scores or conventional statistical analyses that do not ensure measurement invariance across individuals and items. Second, within the domain of literary education—particularly literary psychology—there is a scarcity of research that systematically integrates immersive media with rigorous psychometric validation. This gap is particularly problematic given the inherently interpretive, multidimensional, and subjective nature of literary cognition, which requires measurement approaches capable of capturing latent traits with precision.

In this regard, the Rasch model provides a theoretically grounded and methodologically robust framework for evaluating learning instruments in immersive environments. Unlike classical test theory, Rasch analysis enables the transformation of ordinal data into interval-level measures, ensures parameter invariance, and offers detailed diagnostic information regarding item functioning and respondent behavior. This is particularly relevant in augmented reality–based learning, where multimodal stimuli and interactive features may introduce additional sources of response variability. By applying Rasch measurement, this study not only evaluates the effectiveness of immersive literary learning but also critically examines whether the assessment instrument itself operates validly within such complex learning contexts. Therefore, integrating Rasch analysis into immersive learning research is essential to bridge the gap between pedagogical innovation and measurement validity, ensuring that observed learning gains are both empirically sound and theoretically defensible.

Based on these gaps, this study positions its primary contribution in the psychometric validation of an assessment instrument designed for immersive literary learning environments. Rather than focusing solely on the development of augmented reality–based instructional media, this research emphasizes the importance of ensuring that the measurement of learning outcomes—particularly cognitive engagement—meets rigorous standards of validity and reliability. By applying the Rasch model, this study systematically evaluates the functioning of items, the consistency of participant responses, and the alignment between item difficulty and learner ability within an augmented reality–supported literary psychology context.

In addition to this primary contribution, the study also provides empirical insights into how augmented reality environments influence students' cognitive engagement in literary learning, particularly in tasks involving interpretation, symbolic reasoning, and psychological analysis. Furthermore, the integration of immersive media with Rasch-based measurement offers a comprehensive framework that bridges pedagogical innovation and measurement precision. Therefore, this study contributes not only to the advancement of immersive learning design in literary education but, more importantly, to the establishment of robust measurement practices necessary for evaluating complex learning processes in technology-enhanced environments.

2. METHODS

This study uses a quantitative design with a Rasch model-based measurement approach to evaluate the quality of quiz instruments in the Augmented Reality Psychology Literature (ARPS) application. The Rasch model was chosen because it is able to transform categorical data into interval scales, as well as providing a robust analytical framework to test the suitability of items and participant responses simultaneously (Boone, 2020; Bond & Fox, 2015; Tennant & Conaghan, 2007). This approach allows the identification of the latent structure of participants' abilities while detecting response anomalies that do not conform to probabilistic models (Linacre, 2002; Engelhard, 2013).

The research participants consisted of 52 Indonesian language and literature education students who participated in augmented reality-based literary psychology learning. The selection of participants was carried out purposively, with the criteria of having used the ARPS application fully in the learning session. This sample size is adequate for Rasch analysis on dichotomous instruments, especially in the context of exploration of instrument quality at the development stage (Linacre, 1994; Boone, 2020).

The research instrument is in the form of a digital quiz in the ARPS application which consists of 24 dichotomous questions. The item is designed to measure cognitive engagement on three levels: concept understanding, psychological interpretation of characters, and symbolic inference. The arrangement of items refers to the principles of construct validity and cognitive hierarchy, so that each item has a different level of complexity (Wilson, 2005; Brookhart, 2010). What distinguishes this instrument is its integration with an augmented reality environment, where the stimulus of the question is not only text-based, but also supported by interactive visuals that enrich the context of interpretation (Ibáñez & Delgado-Kloos, 2018; Radianti *et al*., 2020).

The instrument was designed to measure cognitive engagement across three hierarchical levels. First, conceptual understanding refers to the ability to identify and comprehend fundamental literary psychological concepts, such as character traits and narrative structure. Second, interpretative analysis involves the ability to infer psychological states, emotional dynamics, and character motivations based on textual and contextual cues. Third, symbolic and inferential reasoning represents higher-order cognitive processing, where participants are required to interpret abstract meanings, symbolic representations, and implicit psychological conflicts within the narrative. This hierarchical structure reflects increasing cognitive complexity and aligns with higher-order thinking frameworks in literary analysis.

The ARPS (Augmented Reality Psychology Sastra) application was developed as a mobile-based immersive learning platform operating on Android devices using marker-based augmented reality technology. The application was built using Unity 3D integrated with the Vuforia Engine to enable real-time visualization of three-dimensional representations of literary psychological constructs. The minimum hardware requirements included a smartphone with at least 4 GB RAM, a rear camera resolution of 8 MP or higher, and Android OS version 8.0 or above to ensure stable rendering of AR objects.

In terms of interaction design, ARPS employs a user-centered interface that integrates visual markers, interactive 3D objects, and embedded quiz prompts. Users interact with the system by scanning predefined markers, which trigger the appearance of visualized psychological elements such as character emotions, conflict dynamics, and symbolic representations. These elements are manipulable (*e.g.*, rotate, zoom) to facilitate deeper cognitive engagement. The learning scenario consists of a structured sequence: (1) exposure to literary text excerpts, (2) AR-based visualization of psychological constructs, and (3) immediate formative assessment through embedded quiz items.

Each learning session lasted approximately 60–75 minutes, during which participants engaged continuously with the AR environment. The immersive experience supports literary interpretation by transforming abstract psychological dimensions—such as internal conflict and symbolic meaning—into concrete visual representations, thereby reducing interpretive ambiguity and enhancing inferential reasoning processes.

To clarify the context of the instrument, the ARPS application interface is presented in the methods section as evidence of the implementation of the learning media.

Figure 1: ARPS Application Primer.

The main interface of the ARPS application as shown in Figure **1** shows a systematic navigation structure, which facilitates gradual access from the material to evaluation through the quiz feature.

Figure 2: Sample Quiz Items in ARPS.

The figure illustrates how literary psychology questions are embedded within the application, including immediate feedback to support cognitive reinforcement.

The study employed a one-group pre-test–post-test design to examine changes in cognitive engagement following interaction with the ARPS application. Participants first completed a pre-test to assess baseline cognitive performance in literary psychology. Subsequently, they engaged in an augmented reality–based learning session lasting approximately 60–75 minutes. After the intervention, participants completed a post-test using the same instrument calibrated under the Rasch model framework (Boone, 2020; Bond & Fox, 2015; Linacre, 2002).

This design allows for the examination of changes in person ability estimates (logits) before and after exposure to the immersive learning environment. While no control group was included, the within-subject comparison provides initial evidence of the effectiveness of AR-based learning in enhancing cognitive engagement. Future studies are recommended to incorporate comparative designs involving traditional instructional methods to further validate these findings.

3. RESULTS

3.1. Summary of Measurement Statistics

The instrument showed adequate measurement quality with a person reliability of 0.78 and an item

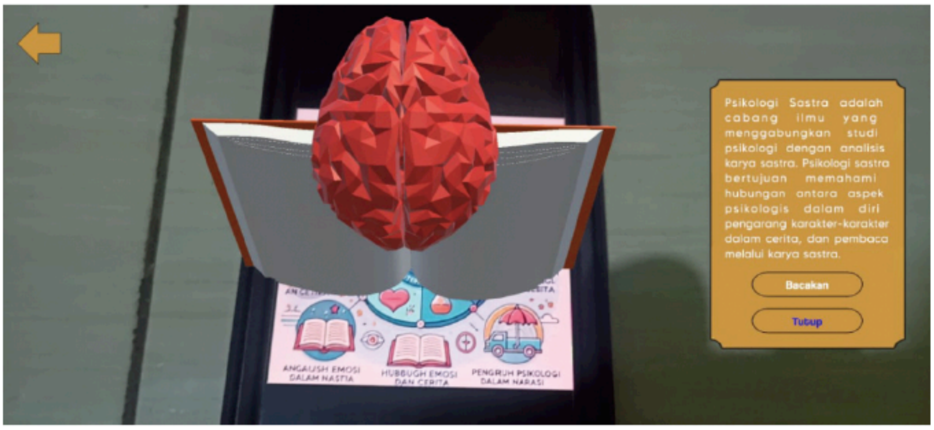

Figure 3: Implementation of Augmented Reality in Learning.

Three-dimensional representations are used to externalize abstract concepts such as character emotions and internal conflicts, enabling more concrete interpretation.

reliability of 0.91, which indicated the consistency of participant responses and the stability of the difficulty level calibration of the item. A person separation value of 1.88 indicates that the instrument is able to distinguish about two to three skill strata, while an item separation of 3.18 confirms the existence of a clear and well-distributed hierarchy of item difficulty along the logit continuum (see Table **1**).

Table 1: Rasch Measurement Summary Statistics

Parameter	Value
Person Reliability	0.78
Item Reliability	0.91
Person Separation	1.88
Item Separation	3.18
Cronbach's Alpha (KR-20)	0.81
Mean Person Ability (logit)	+0.42
Mean Item Difficulty (logit)	0.00

Narrative:

The average ability of participants (+0.42 logits) that is above the average difficulty of the item (0.00 logit) indicates the existence of targeting mismatch, where the instrument is relatively easier for most participants. This condition has the potential to reduce measurement accuracy at high capability levels due to the limitations of items with more complex difficulty levels. In Rasch's framework, this mismatch suggests that the distribution of items has not been fully optimal to capture the variation in capabilities across the board (Bond & Fox, 2015). The implication is that even if the instrument has met measurement reliability and stability standards, the development of item with higher levels of cognitive complexity is needed to improve measurement sensitivity, particularly in the context of augmented reality-based learning that tends to improve participant performance through visual and interactive support.

3.2. Item Fit Analysis

In general, all items show a good fit with the Rasch model, with the MNSQ infit and outfit values in the range of 0.88–1.22. This range indicates that the item is functioning productively for measurements without showing significant distortion. A consistently positive point-measure correlation value (0.31–0.49) confirms that each item validly contributes to the cognitive engagement construct.

Judging from the difficulty distribution (logit), the instrument forms a relatively continuous gradient from easy item (*e.g.* I01–I03; < −0.75 logits) to hard item (*e.g.* I22–I24; > 1.10 logits). This structure reflects an organized cognitive hierarchy, from the introduction of basic concepts to inferential analysis in literary psychology. However, item density in the medium range (approximately −0.30 to 0.90 logits) is more dominant than in the extreme range, which is consistent with the indication of targeting mismatch in Section 3.1.

From a diagnostic perspective, some items are in a borderline misfit condition (*e.g.* I11, I15, I23 with outfits close to 1.20–1.22). Although still within acceptable limits, this pattern indicates an increase in response variability in items with higher cognitive demands. Substantively, this condition tends to be related to the complexity of interpretation, in which participants not only access explicit information, but also perform contextual integration of psychological meaning.

In contrast, some items with an MNSQ value of < 0.90 (*e.g.* I03, I09, I19) show a tendency to overfit,

Table 2: Item Fit Statistics (24 Item Summary)

Item	Logit	Infit MNSQ	Outfit MNSQ	Pt-Measure Corr
I01	-1.12	0.94	0.92	0.41
I02	-0.98	1.02	1.05	0.39
I03	-0.75	0.89	0.87	0.45
I04	-0.60	1.10	1.12	0.36
I05	-0.42	0.97	0.95	0.43
I06	-0.20	1.06	1.08	0.35
I07	0.05	0.92	0.90	0.47
I08	0.18	1.12	1.15	0.33
I09	0.35	0.88	0.86	0.49
I10	0.52	1.07	1.09	0.37
I11	0.70	1.18	1.22	0.31
I12	0.88	0.96	0.98	0.40
I13	1.05	1.09	1.12	0.34
I14	1.20	0.91	0.89	0.46
I15	1.35	1.14	1.18	0.32
I16	1.48	0.98	1.01	0.41
I17	-0.85	0.95	0.93	0.44
I18	-0.30	1.03	1.05	0.38
I19	0.22	0.90	0.88	0.47
I20	0.65	1.11	1.14	0.33
I21	0.90	1.07	1.09	0.36
I22	1.10	0.93	0.91	0.45
I23	1.25	1.15	1.20	0.31
I24	1.40	0.97	0.99	0.40

which indicates an over-consistent response and potential for construct redundancy. Within the framework of instrument development, this condition can reflect the scope of the cognitive aspect that is too narrow or the structural similarity between the items.

The implication is that, while no items need to be eliminated, instrument optimization should be focused on:

a) Enhancement of borderline item to improve clarity of cognitive demands and item quality, and

b) Reduction of redundancy on overfit item to improve measurement efficiency without reducing construct representation.

In the context of augmented reality-based learning, the stability of these fit items indicates that immersive interactions and live feedback within the app do not result in significant response distortions. Thus, the performance of participants on the quiz can be interpreted as a representation of measured cognitive ability, rather than as an artifact of the learning medium.

Evaluation of respondents' suitability to the Rasch model was carried out to ensure that participants' response patterns were consistent with the model's probabilistic expectations, so that the resulting ability estimates could be validly interpreted. This analysis becomes important to identify potential distortions that come from inconsistent responses, such as random guesses or attention fluctuations during the work process.

As summarized in Table **3**, the mean scores of infit (1.01) and MNSQ outfit (1.03) were very close to the ideal value (1.00), which suggests that in general the response patterns of participants followed the structure of the Rasch model. This indicates that the estimated capabilities produced are stable and not affected by significant response noise.

Table 3: Person Fit Summary

Parameter	Value
Mean Infit MNSQ	1.01
Mean Outfit MNSQ	1.03
SD Infit	0.19
SD Outfit	0.23
Person Misfit (>1.5)	3 (5.8%)

Relatively low response variability (SD infit = 0.19; SD outfit = 0.23) indicates that there are no extreme deviations in the answer pattern between respondents. This condition confirms that the participant's interaction with the instrument takes place consistently and structured, with no indication of a dominant random response.

A small number of respondents (5.8%) were identified as misfit (MNSQ outfit > 1.50), indicating a mismatch between individual and model response patterns. In the context of measurement, this condition is generally related to non-cognitive factors, such as inconsistencies in answering, guessing, or variations in the level of engagement during the learning process.

These findings are also consistent with the ability distribution on the Wright map (see Figure **4**), where most respondents show a structured distribution pattern without extreme anomalies. This reinforces that the identified response mismatch is limited and does not interfere with the overall stability of the measurement model. Thus, the overall quality of participants' responses can be categorized as good, so that the results of the ability estimate are suitable for further analysis without significant correction to the data.

3.3. Wright Map (Person–Item Map Interpretation)

Person-item targeting evaluation was carried out to assess the extent to which the distribution of participants' abilities was in line with the distribution of the difficulty level of the items in the instrument. This suitability is an important indicator in the Rasch model, as it determines the level of measurement precision along the capability continuum. As shown in Table **4**, the average ability of the participant is at +0.42 logits, higher than the average difficulty of the item calibrated at 0.00 logits. This difference resulted in a targeting gap of +0.42 logits, indicating that the instrument tended to be relatively easier for most participants.

Table 4: Person–Item Targeting Summary

Parameter	Value
Mean Person Ability (logit)	+0.42
Mean Item Difficulty (logit)	0.00
Targeting Gap	+0.42
Person SD	0.86
Item SD	0.74

A visualization of the distribution of participant abilities and item difficulty at the same logit scale is shown in Figure **5**. The Wright map shows that most participants are distributed at levels above average item difficulty, with higher densities in the medium to high ability range. In contrast, item distribution tends to be concentrated on medium difficulty, with the number of items on high difficulty relatively limited.

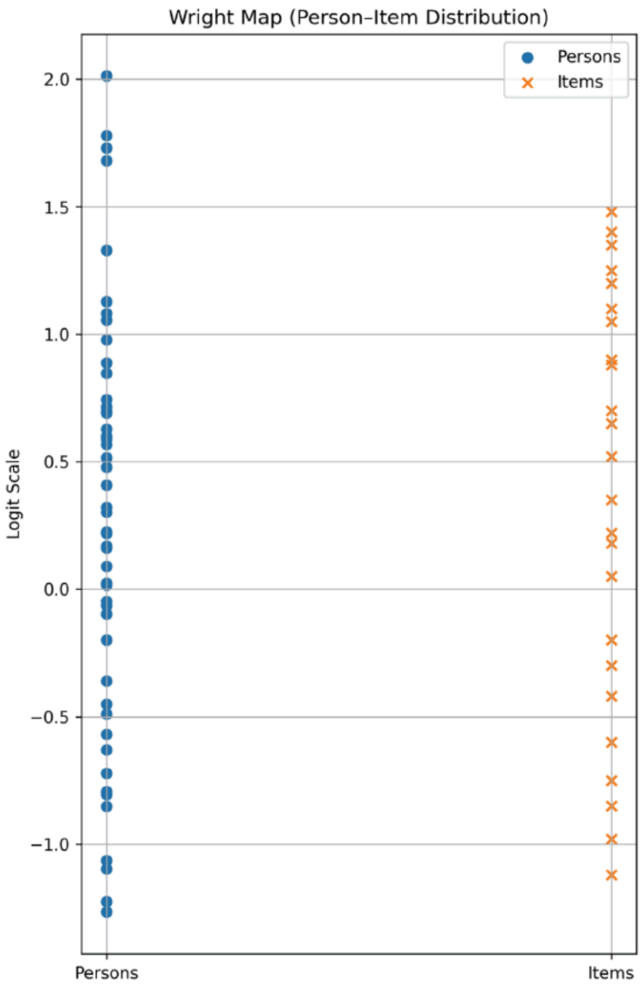

Figure 4: Wright Map displaying the distribution of person abilities (**left**) and item difficulties (**right**) on a common logit scale.

A more detailed inspection of Figure **4** reveals a clear imbalance between the distribution of person abilities and item difficulties. On the left side of the Wright map, most participants are clustered above the mean logit (0.00), indicating that a substantial proportion of respondents possess higher ability levels. In contrast, the right side shows that item difficulties are predominantly concentrated around the medium range, with relatively fewer items located at higher difficulty levels.

This asymmetry suggests that the instrument is insufficiently challenging for participants with higher abilities, resulting in a targeting mismatch. Specifically, the density of persons at higher logits is not matched by a corresponding density of difficult items, which limits the instrument's ability to discriminate among high-performing participants. Consequently, measurement precision at the upper end of the ability spectrum is reduced.

Furthermore, the relatively even spread of items in the lower and middle logit ranges indicates that the instrument performs adequately in distinguishing participants with low to moderate abilities. However, the lack of items beyond +1.0 logits highlights the need for the development of more cognitively demanding items, particularly those requiring deeper inferential reasoning and symbolic interpretation in literary contexts.

4. DISCUSSION

The findings of this study show that the instruments in the Augmented Reality Psychology Literature (ARPS) application have stable measurement quality, as demonstrated by the adequate reliability and suitability of the items and respondents to the Rasch model. In the framework of fundamental measurement, the compatibility between empirical data and probabilistic models indicates that the estimation of ability is invariant and can be validly interpreted (Boone, 2020; Bond & Fox, 2015; Tennant & Conaghan, 2007). The stability of the item fit and person fit found also reinforced that the data were not dominated by random responses or structural distortions, so that the measurement results had a strong psychometric foundation (Linacre, 2002; Wright & Masters, 1982; Engelhard, 2013).

The distribution of the difficulty level of the items that form a continuous gradient shows that the instrument has adequately represented the cognitive hierarchy, from basic comprehension to inferential ability. This structure is in line with the cognitive taxonomic framework that emphasizes the progression from lower-order thinking to higher-order thinking skills (Anderson & Krathwohl, 2001; Brookhart, 2010). In the context of literary psychology, these abilities include character interpretation, emotional dynamics, as well as symbolic analysis that requires the integration of conceptual knowledge and interpretive sensitivity (Eagleton, 2008; Wellek & Warren, 1949). However, indications of overfit in some items indicate the presence of construct redundancy, which has the potential to reduce measurement efficiency if not optimized (Tennant & Conaghan, 2007; Bond & Fox, 2015).

The key finding is that targeting mismatch, where a participant's ability goes beyond an item's difficulty level, has important implications in Rasch's perspective. This discrepancy indicates that the instrument is not yet fully able to distinguish participants at a high level of ability, so the measurement precision at the upper end of the distribution is limited (Wilson, 2005; Boone, 2020). In the measurement literature, alignment between the distribution of ability and difficulty of items is a key prerequisite for producing accurate and sensitive estimates of individual variation (Bond & Fox, 2015; Linacre, 2002).

However, this condition can also be interpreted as a reflection of the effectiveness of augmented reality-based learning used in this study. A number of studies show that AR technology is able to improve engagement and understanding through visual integration and interactivity that enriches the learning experience (Ibáñez & Delgado-Kloos, 2018; Radianti *et al*., 2020; Akçayır & Akçayır, 2017). Immersive learning environments allow participants to process information more efficiently, especially in abstract and interpretive domains (Dunleavy & Dede, 2014; Bacca *et al*., 2014).

From a user experience perspective, the effectiveness of the ARPS application can also be interpreted through the lens of immersion, interactivity, and usability. Immersive learning environments are characterized by a sense of presence, where users feel cognitively and perceptually engaged within the learning space. In this study, the integration of interactive 3D visualizations and real-time feedback contributed to sustained attention and deeper engagement with literary content. The usability of the interface—characterized by intuitive navigation, responsive interaction, and clear visual cues—further supported cognitive processing without introducing excessive cognitive load.

Previous studies in immersive learning and AR-based user experience suggest that the combination of interactivity and multimodal representation enhances both engagement and comprehension (Radianti *et al*., 2020; Makransky & Petersen, 2019; Ibáñez & Delgado-Kloos, 2018). In the context of literary learning, these affordances are particularly important, as they allow abstract interpretive processes to be grounded in perceptual experiences. Therefore, the observed improvement in cognitive engagement is not only a function of content delivery but also a result of well-designed user interaction within the immersive environment.

However, the methodological implications of this condition need to be critically considered. When the instrument is unable to keep up with the increase in participants' abilities, the sensitivity of the measurement becomes limited. Therefore, future development of instruments needs to be focused on increasing the cognitive complexity of items, especially those that demand in-depth inference, symbolic analysis, and integration of contextual meaning (Wilson, 2005; Messick, 1995, Parong *et al*., 2021). This confirms that the validity of measurements is not static, but must be responsive to changes in the learning context and characteristics of participants (Messick, 1995; Kane, 2013).

Overall, the study shows that the integration between augmented reality technology and the Rasch approach not only provides a more comprehensive understanding of learning effectiveness, but also highlights the importance of adapting evaluation instruments in the context of immersive learning. Thus, the contribution of this research lies in strengthening the relationship between pedagogical innovation, cognitive representation, and measurement validity in the domain of technology-based literary psychology.

CONCLUSION

This study demonstrates that the integration of augmented reality in literary psychology learning enhances students' cognitive engagement while maintaining robust measurement quality through Rasch analysis. The acceptable reliability indices, appropriate item and person fit, and the hierarchical structure of item difficulty confirm that the instrument consistently represents the intended cognitive constructs. However, the presence of a targeting mismatch—where participant abilities exceed item difficulty—indicates that while augmented reality effectively facilitates learning performance, it simultaneously requires the development of more cognitively demanding items to improve measurement sensitivity, particularly at higher ability levels. These findings underscore the importance of aligning instructional innovation with measurement precision in immersive learning environments.

From an applied perspective, the study contributes both pedagogically and methodologically. Pedagogically, augmented reality offers a powerful approach to support abstract and interpretive processes in literary learning by transforming complex psychological constructs into interactive visual representations. Methodologically, the use of the Rasch model strengthens the validity and precision of measurement in technology-enhanced contexts. Nevertheless, this study is limited by its implementation within a single institutional setting, which may restrict the generalizability of the findings, and by the absence of a control group, which limits causal interpretation. In addition, the current distribution of item difficulty has not fully captured higher levels of participant ability. Future research should therefore involve more diverse and multi-institutional samples, incorporate controlled experimental designs, and develop higher-complexity items, as well as longitudinal approaches, to better examine the stability and progression of cognitive engagement in immersive learning environments.

REFERENCE

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review. Educational Research Review, 20, 1-11. https://doi.org/10.1016/j.edurev.2016.11.002

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. Educational Technology & Society, 17(4), 133-149.

Boone, W. J. (2020). Rasch analysis for instrument development: Why, when, and how? CBE—Life Sciences Education, 19(4), es9.

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Routledge.

Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

Bruner, J. (1990). Acts of meaning. Harvard University Press.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. Journal of Science Education and Technology, 23, 735-748. https://doi.org/10.1007/978-1-4614-3185-5_59

Eagleton, T. (2008). Literary theory: An introduction. Blackwell.

Engelhard, G. (2013). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. Routledge. https://doi.org/10.4324/9780203073636

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. Computers & Education, 128, 1-13.

Garzón, J., & Acevedo, J. (2021). Meta-analysis of the impact of augmented reality on students' learning gains. Educational Research Review, 33, 100411.

Ibáñez, M. B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. Computers & Education, 123, 109-123. https://doi.org/10.1016/j.compedu.2018.05.002

Kane, M. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1-73. https://doi.org/10.1111/jedm.12000

Linacre, J. M. (1994). Sample size and item calibration stability. Rasch Measurement Transactions, 7(4), 328.

Linacre, J. M. (2002). What do infit and outfit mean-square values mean? Rasch Measurement Transactions, 16(2), 878.

Makransky, G., & Petersen, G. B. (2019). Immersive virtual reality and learning: A meta-analysis. Educational Psychology Review, 31(1), 1-40.

Makransky, G., Petersen, G. B., & Klingenberg, S. (2021). Can immersive virtual reality simulations facilitate learning? A meta-analysis. Computers & Education, 151, 103860.

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2023). Immersive virtual reality and learning: A meta-analysis of immersion and cognitive load effects. Educational Psychology Review, 35(2), 1-27.

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (pp. 43-71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses. American Psychologist, 50(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Paivio, A. (1991). Dual coding theory: Retrospect and current status. Canadian Journal of Psychology, 45(3), 255-287. https://doi.org/10.1037/h0084295

Parong, J., Mayer, R. E., Fiorella, L., MacNamara, A., Homer, B. D., & Plass, J. L. (2021). Learning in immersive virtual reality: Effects of immersion on learning, cognitive load, and transfer. Journal of Educational Psychology, 113(4), 785-800.

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. Personal and Ubiquitous Computing, 18, 1533-1543. https://doi.org/10.1007/s00779-013-0747-y

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education. Computers & Education, 147, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rosenblatt, L. (1994). The reader, the text, the poem: The transactional theory of the literary work. Southern Illinois University Press.

Sweller, J. (2011). Cognitive load theory. Psychology of Learning and Motivation, 55, 37-76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology. Arthritis Care & Research, 57(8), 1358-1362. https://doi.org/10.1002/art.23108

Wellek, R., & Warren, A. (1949). Theory of literature. Harcourt.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Lawrence Erlbaum Associates.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. MESA Press.

https://doi.org/10.65638/2978-8811.2026.02.03