

DIGITAL GAME EXPERIENCE, COLLECTIVE MEMORY AND RAID DESIGN: A MULTI-CRITERIA EVALUATION OF WORLD OF WARCRAFT RAIDS

Burak BORA

Nevşehir Hacı Bektaş Veli Üniversitesi, Türkiye

burakbora@nevsehir.edu.tr

<https://orcid.org/0000-0002-4723-3405>

Atıf	Bora, B. (2026). Digital Game Experience, Collective Memory and Raid Design: A Multi-Criteria Evaluation of World of Warcraft Raids, <i>Yeni Medya Elektronik Dergisi</i> , 10 (3), 457-488.
------	--

ABSTRACT

Massively multiplayer online role-playing games have become persistent cultural environments in which game mechanics, narrative structures, social coordination and collective player memory interact over long periods of time. Within this context, raid content occupies a central position because it combines encounter design, group coordination, progression systems, narrative closure and long-term replay value. Despite the extensive scholarly attention given to online games, virtual worlds and MMORPG player practices, comparative evaluations of raid design have largely remained within fan rankings, informal retrospectives and community discussions. This study addresses this gap by developing a transparent multi-criteria decision-making framework for evaluating selected *World of Warcraft* raids. Twenty raids released between Classic and *Dragonflight* were assessed across ten criteria covering structural scale, encounter diversity, narrative integration, atmosphere, end-boss iconicity, replay appeal, community memory, accessibility burden and design innovation. Criterion weights were calculated using the CRITIC method, and raid performance was ranked through TOPSIS. Entropy weighting, equal weighting, Spearman rank correlation and $\pm 10/\pm 20$ sensitivity analyses were applied to test robustness. The results identify Ulduar as the most robustly superior raid, ranking first under all weighting schemes and all sensitivity scenarios. Karazhan, Icecrown Citadel and The Nighthold form a stable upper-tier group, while mid-ranked raids show greater sensitivity to weighting assumptions. The study contributes to game studies by

operationalising raid design as a multi-dimensional evaluation problem and demonstrates how MCDM methods can complement qualitative and community-based interpretations of digital game experience.

Keywords: *MMORPG; World of Warcraft; Raid Design; Digital Game Experience; Multi-Criteria Decision-Making.*

DİJİTAL OYUN DENEYİMİ, KOLEKTİF BELLEK VE RAID TASARIMI: WORLD OF WARCRAFT RAIDLERİNİN ÇOK KRİTERLİ DEĞERLENDİRİLMESİ

ÖZ

Kitlesel çok oyunculu çevrim içi rol yapma oyunları, oyun mekanikleri, anlatı yapıları, sosyal koordinasyon ve kolektif oyuncu belleğinin uzun dönemli olarak etkileşime girdiği kalıcı dijital kültür ortamlarıdır. Bu bağlamda raid içerikleri; karşılaşma tasarımı, grup koordinasyonu, ilerleme sistemleri, anlatısal kapanış ve tekrar oynanabilirlik boyutlarını bir araya getirdiği için özel bir önem taşımaktadır. Bu çalışma, seçilmiş World of Warcraft raidlerini değerlendirmek için şeffaf bir çok ölçütlü karar verme çerçevesi geliştirmektedir. Classic döneminden Dragonflight dönemine kadar yayımlanan yirmi raid; yapısal ölçek, karşılaşma çeşitliliği, anlatı bütünleşmesi, atmosfer, son boss ikonikliği, tekrar oynanabilirlik, topluluk belleği, erişilebilirlik yükü ve tasarım yeniliği gibi on ölçüt üzerinden değerlendirilmiştir. Ölçüt ağırlıkları CRITIC yöntemiyle hesaplanmış, raid performansları ise TOPSIS yöntemiyle sıralanmıştır. Entropi ağırlıklandırma, eşit ağırlıklandırma, Spearman sıra korelasyonu ve $\pm\%10/\pm\%20$ duyarlılık analizleriyle bulguların sağlamlığı test edilmiştir. Sonuçlar, Ulduar'ın tüm ağırlıklandırma modelleri ve duyarlılık senaryoları altında birinci sırada kaldığını göstermektedir. Çalışma, raid tasarımını çok boyutlu bir değerlendirme problemi olarak operasyonelleştirerek dijital oyun çalışmaları literatürüne katkı sunmaktadır.

Anahtar Kelimeler: *MMORPG; World of Warcraft; Raid Tasarımı; Dijital Oyun Deneyimi; Çok Ölçütlü Karar Verme.*

INTRODUCTION

Digital games are no longer marginal entertainment objects but complex cultural, technological and social systems. Since the early institutionalisation of game studies, scholars have argued that digital games require dedicated analytical frameworks because they combine rule-based systems, representational worlds, player agency and situated experience in ways that cannot be fully explained by earlier media theories alone (Aarseth, 2001; Juul, 2005). Game design research has therefore increasingly treated games as systems of rules, interaction, feedback and experience, rather than merely as texts or narratives. The MDA framework, for instance, conceptualises games through the relationship between mechanics, dynamics and aesthetics, while Sicart's work on game mechanics emphasises the actions available to players within rule-governed possibility spaces (Hunicke et al., 2004; Sicart, 2008). These perspectives are particularly relevant for massively multiplayer online role-playing games (MMORPGs), where player experience emerges not only from individual interaction with rules and fiction, but also from persistent worlds, group coordination, social norms and long-term community practices.

Among MMORPGs, *World of Warcraft* represents one of the most influential and long-lasting cases for examining digital play as a cultural and social practice. Previous studies have analysed *World of Warcraft* in relation to social interaction, player motivation, ethnographic experience, virtual world participation and community formation. Yee's motivational model shows that online game play is structured by multiple motivational components, including achievement, social interaction and immersion, while Ducheneaut et al. demonstrated that social dynamics in MMORPGs are more complex than simple assumptions about online sociability suggest (Yee, 2006; Ducheneaut et al., 2006). Nardi's ethnographic account further positioned *World of Warcraft* as a site of cultural practice and aesthetic experience, while later work on raiding and MMORPG group activity has shown how game design, norms, skills and coordinated actions intersect in collective play (Nardi, 2010; Raudonat & Marsden, 2018).

Raid content is especially important within this broader MMORPG ecology. Raids are not merely high-level instances or boss encounters; they are structured collective experiences that require coordination, preparation, role differentiation, strategic learning and repeated participation. In this sense, raids are simultaneously design artefacts, social events and memory-producing cultural forms. Golub's ethnographic work on *World of Warcraft* raiding, for example, treats raiding as a form of knowledge production in which realism, coordination and expertise are enacted through practice (Golub, 2010). Similarly, studies of

collective memory in virtual worlds suggest that persistent game environments can generate shared histories, group identities and remembered events that outlast individual play sessions (Papargyris & Poulymenakou, 2009). This makes raid design an appropriate object for evaluating how mechanics, narrative, atmosphere, social memory and replay value combine in digital game experience.

Despite this significance, raid evaluation remains underdeveloped as a formal academic problem. In player communities, raids are frequently ranked through forum debates, retrospective videos, tier lists and fan discussions. Such evaluations are valuable because they reflect lived experience and community memory, but they usually lack methodological transparency. Criteria are often implicit, weights are rarely specified and the distinction between design quality, nostalgia, difficulty, narrative importance and replay value is blurred. Academic game studies, by contrast, has produced rich qualitative and theoretical accounts of virtual worlds, player practices and game design, but has paid less attention to the systematic comparative evaluation of individual content units within long-running games. This creates a methodological gap: raid content is widely evaluated in community discourse, yet rarely operationalised as a structured multi-criteria decision problem.

This study responds to that gap by proposing a multi-criteria decision-making model for evaluating *World of Warcraft* raid design. Multi-criteria decision-making (MCDM) is suitable for problems in which alternatives must be assessed across multiple, potentially conflicting criteria. In the present context, a raid may be strong in atmosphere and community memory but weaker in accessibility; another may have high mechanical diversity but limited narrative integration. A single-dimensional ranking would therefore be insufficient. The CRITIC method is used to derive objective criterion weights from the decision matrix by considering both the contrast intensity of each criterion and inter-criterion conflict (Diakoulaki et al., 1995). TOPSIS is then used to rank raids according to their relative closeness to the positive ideal solution and distance from the negative ideal solution (Hwang & Yoon, 1981). Entropy and equal weighting are included as robustness checks because objective weighting methods may produce different weight structures depending on the distributional properties of the decision matrix.

The empirical design evaluates twenty *World of Warcraft* raids released between Classic and *Dragonflight*. The alternatives include historically foundational raids such as Molten Core and Blackwing Lair, culturally iconic raids such as Karazhan, Ulduar and Icecrown Citadel, and more recent examples such as Castle Nathria and Amirdrassil. Each raid is assessed across ten criteria: boss count,

wing count, encounter diversity, narrative integration, atmosphere/visual identity, end-boss iconicity, replay/farming appeal, community memory, accessibility burden and design innovation. These criteria were selected to capture both measurable structural features and interpretive dimensions of digital game experience. Boss count and wing count represent scale and structural breadth, while the remaining criteria capture design, cultural and experiential qualities that are central to raid evaluation but difficult to reduce to a single metric.

The study makes three contributions. First, it translates raid evaluation from informal fan ranking into a transparent decision model in which criteria, directions, weights and rankings are explicitly reported. Second, it extends game studies by demonstrating how MCDM can be used as a complementary method for analysing digital game content. This does not replace interpretive, ethnographic or cultural approaches; rather, it provides an additional layer of structured comparison that can make implicit evaluative assumptions more visible. Third, the study contributes empirically by identifying the raids that perform most consistently under alternative weighting and sensitivity conditions. The results show that Ulduar is not only ranked first in the baseline CRITIC–TOPSIS model but also remains first under Entropy weighting, equal weighting and all $\pm 10/\pm 20$ sensitivity scenarios. This finding suggests that Ulduar’s status as a benchmark raid is not merely an artefact of one weighting scheme, but reflects a balanced profile across design, narrative, replay and memory-related criteria.

Accordingly, the study is guided by the following research questions:

RQ1. How can MMORPG raid design be operationalised as a multi-criteria evaluation problem?

RQ2. Which *World of Warcraft* raids perform best when evaluated across structural, experiential, narrative and community-memory criteria?

RQ3. How robust are the resulting rankings under alternative weighting schemes and sensitivity scenarios?

The remainder of the article is organised as follows. The next section reviews the relevant literature on MMORPGs, raid design, player experience, collective memory and multi-criteria decision-making. The methodology section explains the sample, criteria, coding procedure, CRITIC weighting, TOPSIS ranking and robustness analyses. The results section presents the criterion weights, baseline ranking, Entropy and equal-weight comparisons, Spearman correlations and sensitivity findings. The discussion section interprets the implications of the

results for digital game studies, game design evaluation and the methodological use of MCDM in cultural analysis.

LITERATURE REVIEW

Digital Games as Complex Cultural and Rule-Based Systems

Digital games have been studied as complex artefacts that combine computational rules, designed interaction, fictional worlds and situated player experience. Early game studies scholarship emphasised that games should not be reduced to cinema, literature or other pre-existing media forms, because digital games organise experience through procedural structures, player action and rule-based systems (Aarseth, 2001). This argument helped establish game studies as a field concerned not only with representation but also with interactivity, mechanics, agency and play. In this respect, digital games are both cultural texts and operational systems: they communicate through images, stories and themes, but they also structure behaviour through rules, challenges, feedback and affordances.

Game design literature has similarly conceptualised games as multi-layered systems. Hunicke, LeBlanc and Zubek's MDA framework defines games through the relationship between mechanics, dynamics and aesthetics, thereby linking design elements to player-facing experience (Hunicke et al., 2004). Mechanics refer to the formal components of the game system, dynamics emerge from player interaction with those mechanics, and aesthetics describe the experiential responses generated through play. This framework is relevant to raid design because raids are not evaluated only by their technical structure, such as number of bosses or spatial layout, but also by the experiential patterns produced through coordinated group play, challenge escalation, narrative tension and memorable encounters.

A complementary perspective is offered by Sicart, who defines game mechanics as methods invoked by agents for interacting with the game world (Sicart, 2008). This formulation is useful for analysing raid encounters because raid design is largely built around repeated interaction methods: positioning, resource management, role coordination, timing, spatial awareness, phase transitions and collective execution. Therefore, a raid cannot be evaluated only as a narrative chapter or a visual environment. It is also a configuration of mechanics that produces varying degrees of complexity, accessibility, innovation and replayability.

From this perspective, evaluating raid content requires a framework that can capture both formal and experiential dimensions. Boss count and wing structure indicate scale and breadth, but they are insufficient for assessing the quality or cultural significance of a raid. Narrative integration, atmosphere, end-boss iconicity, encounter diversity, community memory and replay value are also central to how raids are remembered and evaluated. This multi-dimensionality justifies the use of a multi-criteria approach rather than a single indicator of raid quality.

MMORPGS, PLAYER MOTIVATION AND SOCIAL EXPERIENCE

Massively multiplayer online role-playing games are among the most important genres for studying persistent digital worlds, social interaction and long-term player investment. Unlike single-player games, MMORPGs are designed around continuity, progression, social organisation and recurring participation. Their worlds persist beyond individual play sessions, and their communities develop norms, histories, reputations and shared memories over time. *World of Warcraft* has been especially influential in this literature because of its scale, longevity and role in shaping public and academic understandings of MMORPG culture.

Yee's empirical work on online game motivations remains foundational for understanding why players engage with MMORPGs. His model identifies achievement, social and immersion components as major motivational dimensions, showing that online play cannot be explained by a single desire for competition or entertainment (Yee, 2006). Raid content is particularly relevant to all three dimensions. Raids involve achievement through progression and mastery, sociality through coordinated group action, and immersion through narrative, atmosphere and world-building. Therefore, raid evaluation needs to account for the interaction between instrumental, social and experiential forms of value.

Ducheneaut, Yee, Nickell and Moore's study of *World of Warcraft* further complicates the assumption that MMORPGs are simply social spaces. Their analysis of play and grouping patterns shows that players may be "alone together": they inhabit shared worlds and benefit from social presence, but their actual forms of grouping and interaction vary substantially (Ducheneaut et al., 2006). This insight is important for raid analysis because raids represent one of the clearest cases in which MMORPG sociality becomes structurally required. Unlike casual questing or world exploration, raid participation depends on coordination, role differentiation and shared tactical knowledge.

Nardi's ethnographic account of *World of Warcraft* also frames the game as a cultural and aesthetic environment rather than merely a technical platform. Her work highlights everyday practices, player routines, social norms and aesthetic experiences within the game world (Nardi, 2010). Such an approach supports the view that raid content should be treated as part of a broader culture of play. Raids are not only consumable content units; they are recurring sites of learning, social bonding, frustration, prestige, memory formation and identity work.

RAID DESIGN AS COLLECTIVE PLAY AND KNOWLEDGE PRODUCTION

Raid content occupies a distinctive position within MMORPG design because it combines mechanical difficulty, collective execution and narrative climax. Raids typically require players to learn complex encounter mechanics, coordinate roles, manage failure and repetition, and develop shared strategies. In this respect, raiding is not merely a test of individual skill; it is a form of collective performance. The group must transform encounter information into practical knowledge, distribute responsibilities and execute coordinated action under conditions of uncertainty and pressure.

Golub's ethnographic study of *World of Warcraft* raiding is especially relevant here. He argues that raiding involves realism and knowledge production within a virtual world, where players produce meaningful forms of expertise through coordinated practice (Golub, 2010). Raid encounters therefore function as systems that generate practical knowledge: players learn what matters, what must be ignored, how to coordinate attention and how to translate mechanical patterns into successful collective action. This makes raid design suitable for evaluation not only in terms of difficulty but also in terms of how effectively it organises collective learning and mastery.

The collective nature of raiding also means that raid quality is not reducible to immediate enjoyment. Some raids become culturally important because they mark historical moments in the game's development, introduce distinctive mechanics, conclude major narrative arcs or create memorable group experiences. Others may be mechanically demanding but less fondly remembered because of excessive accessibility burden, repetition or weak narrative integration. The evaluation of raid design therefore requires attention to both design-internal qualities and community-level reception.

In this study, criteria such as encounter diversity, end-boss iconicity, replay/farming appeal, accessibility burden and design innovation are used to

operationalise different dimensions of raid quality. These dimensions reflect the idea that raids are simultaneously design objects, social coordination systems and memory-producing events. A raid may be structurally large but experientially weak, or mechanically compact but culturally significant. This makes raid ranking a multi-criteria problem rather than a simple popularity contest.

COLLECTIVE MEMORY, NOSTALGIA AND COMMUNITY EVALUATION IN VIRTUAL WORLDS

Persistent online game worlds produce collective memory because players repeatedly inhabit the same environments, narrate shared events and attach meaning to past experiences. Papargyris and Poulymenakou argue that virtual game worlds can support collective memory through shared practices, histories, identities and persistent social structures (Papargyris & Poulymenakou, 2009). This concept is highly relevant to MMORPG raids because raids often become remembered as milestones: the first clear of a difficult boss, the final confrontation with a major antagonist, the acquisition of rare rewards or the experience of progressing with a guild over time.

Community memory is particularly important in a long-running game such as *World of Warcraft*. Because the game has existed across multiple expansions, technical systems, player generations and design philosophies, raid evaluation is shaped by both historical context and retrospective interpretation. Older raids may benefit from nostalgia and historical importance, while newer raids may benefit from refined encounter design and modern production values. Conversely, older raids may score lower on mechanical diversity or accessibility by contemporary standards, while newer raids may have weaker long-term memory because the community has had less time to retrospectively evaluate them.

This tension creates an important methodological challenge. Informal community rankings often merge several different evaluative dimensions: nostalgia, difficulty, personal memories, guild experiences, lore importance, visual design and reward incentives. While these rankings are meaningful cultural artefacts, they do not usually specify the criteria or weights behind the ranking. As a result, they are difficult to compare or reproduce. A structured decision model can make these assumptions more explicit by separating dimensions such as narrative integration, atmosphere, community memory and replay value.

In the present study, “community memory” is treated as one criterion among several rather than as the sole indicator of quality. This is important because

collective memory can enrich evaluation, but it can also introduce bias. A raid may be remembered strongly because it was historically early, narratively important or associated with a large player population, even if its encounter design is less complex by later standards. Conversely, a mechanically strong raid may not yet have accumulated the same level of collective memory. Including community memory as one criterion allows the model to recognise cultural significance without allowing it to dominate the entire evaluation.

EVALUATING DIGITAL GAME CONTENT THROUGH MULTI-CRITERIA DECISION-MAKING

Although game studies has generated extensive theoretical, ethnographic and qualitative work, structured quantitative evaluation of individual game content units remains less developed. Most formal evaluation studies in games focus on usability, player experience, engagement, educational outcomes, monetisation or platform-level comparisons. However, content units within games—such as levels, dungeons, raids, quests or expansions—also involve multiple evaluative dimensions. A raid can be evaluated through its structural scale, mechanical richness, narrative role, aesthetic identity, social coordination requirements, reward structure and long-term cultural status. These dimensions are not always aligned, which creates a classic multi-criteria decision-making problem.

MCDM methods are designed for decision contexts in which alternatives must be compared across multiple criteria that may differ in scale, direction and importance. The CRITIC method is especially useful when the analyst aims to derive objective weights from the decision matrix rather than rely entirely on subjective expert judgement. Diakoulaki, Mavrotas and Papayannakis introduced CRITIC as an objective weighting method based on two informational properties: the contrast intensity of each criterion and the conflict between criteria (Diakoulaki et al., 1995). In this study, CRITIC is appropriate because the criteria include both structural and interpretive dimensions, and the purpose is to identify which criteria provide the strongest discriminating power within the coded matrix.

TOPSIS is one of the most widely used MCDM ranking methods. Developed by Hwang and Yoon, TOPSIS ranks alternatives according to their relative closeness to the positive ideal solution and distance from the negative ideal solution (Hwang & Yoon, 1981). This logic is suitable for raid evaluation because the best raid is not expected to maximise one criterion alone. Instead, it should approximate a balanced ideal: high design innovation, strong narrative

integration, memorable encounters, replay appeal and community significance, while avoiding excessive accessibility burden.

Entropy weighting and equal weighting are commonly used as robustness checks in MCDM studies. Entropy weighting is based on information theory and assigns higher importance to criteria that provide greater differentiation among alternatives. Equal weighting, by contrast, assumes that all criteria are equally important and is useful as a neutral benchmark. Comparing CRITIC, Entropy and equal-weight TOPSIS rankings allows researchers to assess whether the final ranking is dependent on a particular weighting model or remains stable across alternative assumptions.

Sensitivity analysis is also important in MCDM because rankings may change when criterion weights are modified. Recent reviews of MCDM sensitivity analysis emphasise that testing the vulnerability of results to changes in data or weights provides decision makers with information about the stability of conclusions (Więckowski et al., 2023). In the present study, $\pm 10\%$ and $\pm 20\%$ weight perturbations are used to examine whether the ranking of raids changes under moderate shifts in criterion importance. Spearman rank correlation is then used to evaluate the consistency of rankings across models and scenarios.

RESEARCH GAP AND POSITIONING OF THE PRESENT STUDY

The existing literature provides strong foundations for understanding digital games as rule-based systems, MMORPGs as social and cultural environments, and raids as collective practices. However, there remains a gap between qualitative interpretations of MMORPG culture and the informal evaluative practices found in player communities. Academic studies have examined player motivation, social interaction, virtual world culture and raiding practices, but they have rarely transformed raid evaluation into a transparent, reproducible and multi-criteria model.

This study addresses that gap by treating *World of Warcraft* raids as alternatives in a structured decision problem. The proposed framework does not claim to replace interpretive game analysis or community discourse. Rather, it complements them by making evaluative assumptions explicit. Instead of asking which raid is “best” in an undefined sense, the study asks how raids perform across multiple criteria that capture structural, experiential, narrative and community-memory dimensions. This approach allows raid evaluation to move from implicit preference ranking toward a more transparent analytical framework.

The contribution is therefore both substantive and methodological. Substantively, the study provides a comparative evaluation of twenty *World of Warcraft* raids and identifies the most robustly high-performing alternatives. Methodologically, it demonstrates how CRITIC–TOPSIS, supported by Entropy weighting, equal weighting, Spearman correlation and sensitivity analysis, can be applied to digital game content evaluation. In doing so, it opens a possible route for future studies of expansions, dungeons, quests, game modes, monetisation designs or other modular content forms in digital games.

METHODOLOGY

Research Design

This study employs a quantitative multi-criteria decision-making design to evaluate selected *World of Warcraft* raids across structural, experiential, narrative and community-memory dimensions. The methodological aim is not to determine the “best” raid through subjective preference alone, but to construct a transparent and reproducible decision model in which alternatives, criteria, criterion directions, weights and rankings are explicitly reported. This approach is appropriate because raid quality is not a single-dimensional construct. A raid may be mechanically innovative but difficult to access; another may have strong narrative integration but limited replay appeal. Such trade-offs make raid evaluation a multi-criteria decision problem.

The study combines objective and coded indicators. Objective indicators include structural features such as boss count and wing/section count. Coded indicators capture interpretive dimensions such as encounter diversity, atmosphere, end-boss iconicity, community memory and design innovation. The weighting stage was conducted using the CRITIC method, which determines objective criterion weights based on contrast intensity and inter-criterion conflict. This is useful when direct subjective weighting is either difficult or undesirable. The ranking stage was conducted using TOPSIS, a classical multi-attribute decision-making method that ranks alternatives according to their relative closeness to an ideal solution.

SELECTION OF ALTERNATIVES

The alternatives consist of twenty *World of Warcraft* raids released between Classic and *Dragonflight*. The sample was designed to include raids from

different periods of the game’s development rather than focusing only on recent content. This decision allows the model to compare historically foundational raids, culturally iconic raids, mechanically complex raids and more recent examples of raid design.

The final set of alternatives is presented in Table 1. The sample includes early raids such as Molten Core, Blackwing Lair and Temple of Ahn’Qiraj; highly iconic raids such as Karazhan, Ulduar and Icecrown Citadel; expansion-ending raids such as Dragon Soul, Hellfire Citadel and Antorus; and recent examples such as Castle Nathria and Amirdrassil. Very recent raids released after *Dragonflight* were excluded to reduce recency bias and to ensure that the evaluated alternatives had sufficient community-discourse visibility.

Table 1.

Alternatives Included in the Study

Code	Raid	Expansion
A01	Molten Core	Classic
A02	Blackwing Lair	Classic
A03	Temple of Ahn’Qiraj	Classic
A04	Karazhan	The Burning Crusade
A05	Black Temple	The Burning Crusade
A06	Sunwell Plateau	The Burning Crusade
A07	Ulduar	Wrath of the Lich King
A08	Icecrown Citadel	Wrath of the Lich King
A09	Firelands	Cataclysm
A10	Dragon Soul	Cataclysm
A11	Throne of Thunder	Mists of Pandaria
A12	Siege of Orgrimmar	Mists of Pandaria
A13	Blackrock Foundry	Warlords of Draenor
A14	Hellfire Citadel	Warlords of Draenor
A15	The Nighthold	Legion
A16	Tomb of Sargeras	Legion
A17	Antorus, the Burning Throne	Legion
A18	Battle of Dazar’alor	Battle for Azeroth
A19	Castle Nathria	Shadowlands
A20	Amirdrassil, the Dream’s Hope	Dragonflight

CRITERIA AND OPERATIONAL DEFINITIONS

Ten criteria were used to evaluate the raid alternatives. The criteria were selected to represent both structural and experiential dimensions of raid design. Two criteria, boss count and wing count, were treated as directly measurable structural

indicators. The remaining criteria were coded on a five-point scale according to a predefined coding guide.

Nine criteria were defined as benefit criteria, meaning that higher values indicate better performance. One criterion, accessibility burden, was defined as a cost criterion, meaning that lower values indicate better performance. This distinction is important because TOPSIS requires the analyst to define the direction of preference for each criterion.

Table 2.

Evaluation Criteria

Code	Criterion	Direction	Operational definition
C1	Boss Count	Benefit	Number of boss encounters included in the raid
C2	Wing / Section Count	Benefit	Structural breadth of the raid, based on major wings, sections or progression areas
C3	Encounter Diversity	Benefit	Variety of mechanics, encounter formats, role demands and phase structures
C4	Narrative Integration	Benefit	Degree to which the raid is integrated into the expansion's main storyline, lore and narrative closure
C5	Atmosphere / Visual Identity	Benefit	Distinctiveness of art direction, spatial identity, music, environmental design and thematic coherence
C6	End-Boss Iconicity	Benefit	Cultural and narrative significance of the final boss encounter
C7	Replay / Farming Appeal	Benefit	Long-term replay value, including mounts, transmogs, achievements, nostalgia runs and farming incentives
C8	Community Memory	Benefit	Strength of the raid's long-term presence in player discourse, rankings and retrospective evaluations
C9	Accessibility Burden	Cost	Degree to which the raid is associated with excessive difficulty, logistical friction, attunement burden or accessibility barriers
C10	Design Innovation	Benefit	Degree to which the raid introduced, refined or popularised distinctive design ideas or encounter structures

The inclusion of both benefit and cost criteria allows the model to avoid equating complexity with quality. A raid may be highly complex or historically prestigious but still impose an excessive accessibility burden. Therefore, C9 was coded as a cost criterion to account for the negative effect of excessive friction on overall evaluative performance.

CODING PROCEDURE

The decision matrix was constructed in two stages. First, objective structural data were entered for C1 and C2. Boss count was coded as the number of boss encounters in each raid. Wing or section count was coded according to the major structural divisions of the raid. These indicators provide a basic measure of scale and structural breadth.

Second, C3–C10 were coded using a five-point ordinal scale. A score of 1 indicates very low performance on the relevant criterion, while a score of 5 indicates very high performance. The coding process was based on a structured desk review of raid characteristics, retrospective community evaluations, design features, narrative role and long-term player discourse. To improve internal consistency, the same operational definitions were applied across all alternatives.

Table 3.

Coding Scale For C3–C10

Score	Interpretation
1	Very low presence or very weak performance on the criterion
2	Low presence or below-average performance
3	Moderate or mixed performance
4	Strong performance
5	Very strong, iconic or benchmark-level performance

For the cost criterion C9, the interpretation was reversed in the TOPSIS stage. A higher raw C9 score indicates a greater accessibility burden, which is undesirable. Therefore, C9 was treated as a cost criterion during ideal-solution construction. This means that the positive ideal solution used the minimum C9 value, while the negative ideal solution used the maximum C9 value.

Because several criteria involve interpretive judgement, the coding should be understood as a structured expert desk-coding procedure rather than an automated measurement process. This is consistent with the exploratory purpose of the study: to demonstrate how raid design can be transformed into a transparent multi-criteria evaluation framework. The coded matrix also enables future replication, revision or extension by additional coders, expert panels or player-survey data.

CRITIC WEIGHTING PROCEDURE

The CRITIC method was used to calculate objective criterion weights. CRITIC is based on the idea that a criterion is more informative when it displays high contrast among alternatives and when it is not redundant with other criteria. In

other words, a criterion receives a higher weight when it both differentiates alternatives strongly and contributes information that is not already captured by other criteria. Diakoulaki et al. introduced CRITIC as a method for determining objective weights in multiple-criteria problems by using contrast intensity and conflicting character among criteria.

The CRITIC procedure was implemented as follows.

First, the decision matrix was normalised. For benefit criteria, the normalised value was calculated as:

$$r_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

For the cost criterion, C9 Accessibility Burden, the normalised value was calculated as:

$$r_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)}$$

where x_{ij} denotes the original score of alternative i on criterion j , and r_{ij} denotes the normalised value.

Second, the standard deviation of each normalised criterion was calculated to represent contrast intensity. Third, the correlation matrix among criteria was calculated. The information content of each criterion was then computed as:

$$C_j = \sigma_j \sum_{k=1}^n (1 - \rho_{jk})$$

where σ_j is the standard deviation of criterion j , and ρ_{jk} is the correlation coefficient between criteria j and k . Finally, criterion weights were obtained by dividing each criterion's information content by the total information content:

$$w_j = \frac{C_j}{\sum_{j=1}^n C_j}$$

This procedure produced the CRITIC weights used in the baseline TOPSIS analysis.

TOPSIS RANKING PROCEDURE

After the CRITIC weights were obtained, the alternatives were ranked using TOPSIS. TOPSIS is based on the principle that the best alternative should be closest to the positive ideal solution and farthest from the negative ideal solution. The method is widely used in multi-attribute decision-making because it provides a clear distance-based ranking logic and produces an interpretable closeness coefficient for each alternative.

The TOPSIS procedure was implemented in six steps.

First, the original decision matrix was vector-normalised:

$$v_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}$$

Second, the weighted normalised matrix was constructed by multiplying each normalised criterion value by its CRITIC weight:

$$y_{ij} = w_j v_{ij}$$

Third, the positive ideal solution and negative ideal solution were determined. For benefit criteria, the positive ideal solution is the maximum weighted value and the negative ideal solution is the minimum weighted value. For the cost criterion, C9 Accessibility Burden, this logic is reversed.

Fourth, the Euclidean distance of each alternative from the positive ideal solution was calculated:

$$D_i^+ = \sqrt{\sum_{j=1}^n (y_{ij} - y_j^+)^2}$$

Fifth, the Euclidean distance from the negative ideal solution was calculated:

$$D_i^- = \sqrt{\sum_{j=1}^n (y_{ij} - y_j^-)^2}$$

Finally, the relative closeness coefficient was calculated as:

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

Alternatives with higher CC_i values were ranked higher. The closeness coefficient therefore represents each raid's relative proximity to the ideal multi-criteria profile.

ROBUSTNESS CHECKS: ENTROPY AND EQUAL WEIGHTING

To examine whether the baseline ranking was dependent on the CRITIC weighting structure, two additional weighting approaches were applied: Entropy weighting and equal weighting. Entropy weighting is an objective weighting approach based on information differentiation across alternatives. Criteria that display greater variation provide more information and therefore receive greater weight. Entropy-based weighting is frequently used in MCDM as an alternative objective weighting procedure.

Equal weighting was used as a neutral benchmark. Under equal weighting, each of the ten criteria receives a weight of 0.10. Although equal weighting does not reflect differences in criterion informativeness, it is useful for testing whether the main ranking conclusions depend heavily on a specific weighting method.

The TOPSIS procedure was repeated under all three weighting schemes:

1. CRITIC weighting
2. Entropy weighting
3. Equal weighting

The resulting rankings were then compared to assess the stability of the model.

SPEARMAN RANK CORRELATION

Spearman rank correlation was used to measure the consistency among the rankings generated by CRITIC, Entropy and equal weighting. Spearman's rank correlation is appropriate when the objective is to compare ordinal rankings rather than raw numerical values. Spearman's original work established rank-based association as a method for measuring correspondence between ordered variables.

The Spearman coefficient was calculated as:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

where d_i denotes the difference between the ranks of alternative i in two ranking lists, and n denotes the number of alternatives. A coefficient close to 1 indicates strong agreement between rankings, while a coefficient close to 0 indicates weak rank-order consistency.

In this study, Spearman coefficients were calculated for the following pairs:

- CRITIC ranking vs. Entropy ranking
- CRITIC ranking vs. Equal-weight ranking
- Entropy ranking vs. Equal-weight ranking

This analysis provides a direct test of whether the ranking pattern remains consistent across alternative weighting assumptions.

SENSITIVITY ANALYSIS

A one-way sensitivity analysis was conducted to examine whether the ranking results were vulnerable to changes in criterion weights. Sensitivity analysis is an important component of multi-criteria decision analysis because it helps evaluate the stability of ranking outcomes and identify which parameters have the greatest influence on the results. Recent reviews emphasise that sensitivity analysis improves the credibility of MCDA studies by revealing how changes in input data or weights affect final decisions.

Each CRITIC weight was adjusted separately by four perturbation levels:

- -20%
- -10%
- +10%
- +20%

After each adjustment, the remaining weights were proportionally renormalised so that the total weight remained equal to 1. This produced 40 sensitivity scenarios in total, calculated as 10 criteria $\times$ 4 perturbation levels. TOPSIS rankings were recalculated for each scenario.

The sensitivity analysis was evaluated using three indicators:

1. Whether the top-ranked raid changed
2. The maximum rank movement across alternatives

3. The Spearman correlation between each scenario ranking and the baseline CRITIC–TOPSIS ranking

This procedure allows the study to assess whether the main findings are robust or dependent on narrow weight assumptions.

METHODOLOGICAL LIMITATIONS

The methodology has three limitations. First, several criteria rely on structured interpretive coding rather than purely objective measurement. Although the coding guide improves transparency, future studies could strengthen reliability by using multiple independent coders, inter-coder agreement statistics or expert-panel validation. Second, community memory and replay appeal are difficult to measure fully because they are shaped by nostalgia, player cohort, expansion context and evolving game systems. Third, the sample is limited to twenty raids and does not include all *World of Warcraft* raid content.

These limitations do not invalidate the model, but they define its scope. The study should be read as an exploratory and transparent MCDM framework for evaluating raid design rather than a final universal ranking of all *World of Warcraft* raids. Its main methodological contribution is to demonstrate how digital game content can be operationalised as a structured decision problem and subjected to weighting, ranking, robustness and sensitivity analyses.

RESULTS

Criterion Weights Obtained Through Critic

The CRITIC weighting procedure produced a relatively balanced but not uniform distribution of criterion importance. The highest weight was assigned to C9 Accessibility Burden with a weight of 0.1368, indicating that the degree of accessibility-related complexity provided a strong source of contrast among the evaluated raids. This was followed by C6 End-Boss Iconicity with 0.1175, C7 Replay / Farming Appeal with 0.1109, and C10 Design Innovation with 0.1101. These results suggest that the most discriminating dimensions in the dataset were not merely quantitative structural indicators, such as boss count, but rather criteria associated with memorability, replay value, design distinctiveness, and accessibility-related burden.

By contrast, the lowest weights were obtained for C3 Encounter Diversity with 0.0741 and C8 Community Memory with 0.0782. This does not imply that these

criteria are substantively unimportant; rather, within the coded decision matrix, they provided comparatively lower contrast or lower conflict with other criteria. In other words, several raids received similarly high scores on these dimensions, reducing their objective discriminating power in the CRITIC framework.

Table 4.

CRITIC Criterion Weights

Criterion	Direction	CRITIC weight
C9 Accessibility Burden	Cost	0.1368
C6 End-Boss Iconicity	Benefit	0.1175
C7 Replay / Farming Appeal	Benefit	0.1109
C10 Design Innovation	Benefit	0.1101
C4 Narrative Integration	Benefit	0.0975
C2 Wing Count	Benefit	0.0956
C5 Atmosphere / Visual Identity	Benefit	0.0945
C1 Boss Count	Benefit	0.0849
C8 Community Memory	Benefit	0.0782
C3 Encounter Diversity	Benefit	0.0741

Overall, the weighting structure indicates that the evaluation model privileges raids that combine strong design identity, memorable final encounters, replay incentives, and manageable accessibility costs. The relatively high weight of the cost criterion, C9, is particularly important because it means that raids are not rewarded only for complexity or scale; excessive accessibility burden can reduce their overall performance.

BASELINE CRITIC–TOPSIS RANKING

The baseline CRITIC–TOPSIS results are reported in Table 5. The highest-ranked raid was Ulduar, with a TOPSIS closeness coefficient of 0.8111. This result is consistent with Ulduar’s strong profile across several high-impact criteria, particularly design innovation, encounter structure, atmosphere, and end-boss memorability. Its distance from the positive ideal solution was low, while its distance from the negative ideal solution was comparatively high, producing a clear first-place result.

The second and third positions were occupied by Karazhan and Icecrown Citadel, with closeness coefficients of 0.7396 and 0.7366, respectively. The very small difference between these two alternatives indicates that they represent a near-tie under the baseline weighting structure. Karazhan performs strongly because of its distinctive atmosphere, structural originality, and long-term cultural memory, whereas Icecrown Citadel benefits from strong narrative integration and the iconic status of the Lich King as an end-boss encounter.

The next group consists of The Nighthold, Antorus, the Burning Throne, Castle Nathria, and Firelands. These raids form an upper-middle cluster, with closeness coefficients ranging from 0.7086 to 0.6521. Their positions suggest strong but more uneven profiles: for example, The Nighthold benefits from narrative and aesthetic coherence, while Antorus is supported by its end-expansion significance and high-stakes narrative closure.

Table 5.

Baseline CRITIC–TOPSIS Ranking

Rank	Raid	Expansion	TOPSIS CC
1	Ulduar	Wrath of the Lich King	0.8111
2	Karazhan	The Burning Crusade	0.7396
3	Icecrown Citadel	Wrath of the Lich King	0.7366
4	The Nighthold	Legion	0.7086
5	Antorus, the Burning Throne	Legion	0.6749
6	Castle Nathria	Shadowlands	0.6638
7	Firelands	Cataclysm	0.6521
8	Battle of Dazar'alor	Battle for Azeroth	0.6162
9	Amirdrassil, the Dream's Hope	Dragonflight	0.6136
10	Throne of Thunder	Mists of Pandaria	0.6012
11	Blackrock Foundry	Warlords of Draenor	0.5892
12	Siege of Orgrimmar	Mists of Pandaria	0.5707
13	Black Temple	The Burning Crusade	0.5670
14	Hellfire Citadel	Warlords of Draenor	0.5014
15	Molten Core	Classic	0.4990
16	Dragon Soul	Cataclysm	0.4944
17	Tomb of Sargeras	Legion	0.4532
18	Temple of Ahn'Qiraj	Classic	0.4196
19	Blackwing Lair	Classic	0.3974
20	Sunwell Plateau	The Burning Crusade	0.3183

The lower part of the ranking is composed mainly of raids that either have a narrower structural profile, weaker accessibility performance, or more uneven scores across the qualitative design criteria. Sunwell Plateau ranked last with a closeness coefficient of 0.3183. This should not be interpreted as indicating a lack of historical importance; rather, within this multi-criteria model, its compact structure and higher accessibility burden limited its relative performance compared with raids that combined broader design scope, stronger replay value, and more balanced criterion scores.

ROBUSTNESS CHECK WITH ENTROPY AND EQUAL WEIGHTING

To evaluate the robustness of the baseline CRITIC–TOPSIS ranking, two additional weighting schemes were applied: Entropy weighting and equal weighting. The robustness check confirms that the overall rank order is highly stable. The most important finding is that Ulduar remained the first-ranked raid under all three weighting schemes. This indicates that its leading position is not an artefact of the CRITIC weighting structure.

The second and third positions were more sensitive. Under CRITIC and Entropy weighting, Karazhan ranked second and Icecrown Citadel ranked third. Under equal weighting, however, Icecrown Citadel moved to second place and Karazhan moved to third. This suggests that the difference between these two raids is methodologically small and should be interpreted as a competitive upper-tier grouping rather than a strict substantive separation.

Table 6.

Comparison Of TOPSIS Rankings Under Alternative Weighting Schemes

Raid	CRITIC rank	Entropy rank	Equal-weight rank
Ulduar	1	1	1
Karazhan	2	2	3
Icecrown Citadel	3	3	2
The Nighthold	4	4	4
Antorus, the Burning Throne	5	5	7
Castle Nathria	6	8	5
Firelands	7	11	12
Battle of Dazar'alor	8	14	9
Amirdrassil, the Dream's Hope	9	13	13
Throne of Thunder	10	6	6
Blackrock Foundry	11	12	11
Siege of Orgrimmar	12	10	8
Black Temple	13	7	10
Hellfire Citadel	14	9	14
Molten Core	15	17	17
Dragon Soul	16	19	18
Tomb of Sargeras	17	15	15
Temple of Ahn'Qiraj	18	16	16
Blackwing Lair	19	20	20
Sunwell Plateau	20	18	19

The largest movements occurred in the middle of the ranking. Black Temple, for example, improved from 13th under CRITIC to 7th under Entropy, while Battle

of Dazar’alor declined from 8th under CRITIC to 14th under Entropy. These changes indicate that mid-table alternatives are more sensitive to weighting assumptions, especially when their performance profiles are less uniformly strong across the criteria.

SPEARMAN RANK CORRELATION

Spearman rank correlation coefficients were calculated to assess the consistency of the rankings produced by the three weighting schemes. The results show high rank-order agreement across all models. The correlation between CRITIC and Entropy rankings was 0.865, while the correlation between CRITIC and equal weighting was 0.919. The strongest association was observed between Entropy and equal weighting, with a coefficient of 0.938.

Table 7.

Spearman Rank Correlations Between Weighting Schemes

Weighting scheme	CRITIC	Entropy	Equal weight
CRITIC	1.000	0.865	0.919
Entropy	0.865	1.000	0.938
Equal weight	0.919	0.938	1.000

All coefficients exceed 0.86, indicating strong consistency among the ranking models. Therefore, although some alternatives change positions in the middle of the table, the general ranking logic remains stable. The robustness results support the validity of the baseline CRITIC–TOPSIS model, particularly for identifying the leading alternatives.

SENSITIVITY ANALYSIS

A one-way sensitivity analysis was conducted by adjusting each CRITIC criterion weight by $\pm 10\%$ and $\pm 20\%$, resulting in 40 sensitivity scenarios. In every scenario, the adjusted criterion weight was modified while the remaining weights were proportionally renormalized. The purpose of this analysis was to test whether moderate changes in the weighting structure would alter the main ranking conclusions.

The sensitivity results confirm a high level of stability. Ulduar remained the top-ranked alternative in all 40 scenarios. The minimum Spearman correlation between the baseline CRITIC ranking and any sensitivity scenario was 0.970, indicating that even the most disruptive scenario preserved a very similar rank

order. The maximum observed rank movement was four positions, and such movements were concentrated in the middle of the ranking rather than at the top.

Table 8.

Summary Of Sensitivity Analysis

Indicator	Result
Number of scenarios	40
Top-ranked raid in baseline	Ulduar
Top-1 stability	Ulduar in 40/40 scenarios
Minimum Spearman correlation with baseline	0.970
Maximum rank movement	4 positions
Most disruptive scenario	C9 Accessibility Burden –20%

The most disruptive scenario was the 20% decrease in C9 Accessibility Burden, which produced the lowest Spearman correlation with the baseline ranking. This is expected because C9 had the highest CRITIC weight and was the only cost-oriented criterion in the model. Changes in this criterion therefore had a relatively stronger effect on the final ranking. However, even under this scenario, Ulduar retained the first position, and the main top-tier structure remained largely intact.

OVERALL INTERPRETATION OF THE RESULTS

The findings indicate that Ulduar is the most balanced and robustly superior raid in the evaluated sample. Its first-place position is preserved across CRITIC, Entropy, equal weighting, and all sensitivity scenarios. This makes Ulduar not only the baseline winner but also the most methodologically stable alternative in the model.

A second upper-tier group consists of Karazhan, Icecrown Citadel, and The Nighthold. These raids consistently appear near the top of the ranking, although Karazhan and Icecrown Citadel exchange second and third positions depending on the weighting scheme. This suggests that their relative ordering should be interpreted cautiously, while their shared status as top-tier raids is strongly supported.

The middle of the ranking is more fluid. Alternatives such as Throne of Thunder, Black Temple, Siege of Orgrimmar, Firelands, and Battle of Dazar'alor show greater sensitivity to the weighting method. This reflects the fact that these raids possess strong but uneven profiles: they may perform very well on particular dimensions, such as narrative significance or community memory, but less strongly on accessibility, replay value, or structural breadth.

Finally, the lower-ranked raids should not be interpreted as historically insignificant. Several lower-ranked alternatives, including Molten Core, Blackwing Lair, and Sunwell Plateau, are important in the historical development of World of Warcraft raiding. However, the model evaluates them comparatively across ten criteria, and their lower positions reflect limitations in their overall multi-criteria balance rather than a lack of cultural or historical relevance.

DISCUSSION

This study set out to evaluate selected *World of Warcraft* raids through a structured multi-criteria decision-making framework. The findings show that raid design can be operationalised as a multi-dimensional evaluation problem in which structural scale, encounter design, narrative integration, aesthetic identity, replay value, community memory and accessibility burden interact with one another. Rather than treating raid quality as a purely subjective or fan-based preference, the proposed model makes the evaluative logic explicit through criteria, criterion directions, objective weights, TOPSIS closeness coefficients and robustness checks.

The most important empirical finding is the robust first-place position of Ulduar. Ulduar ranked first not only in the baseline CRITIC–TOPSIS model but also under Entropy weighting, equal weighting and all $\pm 10/\pm 20$ sensitivity scenarios. This result suggests that Ulduar’s status as a benchmark raid is not dependent on one particular weighting assumption. Instead, it reflects a balanced performance profile across several dimensions: encounter diversity, narrative integration, atmosphere, design innovation, end-boss memorability and manageable accessibility burden. In this respect, the results support the idea that the strongest raid designs are not necessarily those with the largest number of bosses or the highest level of mechanical difficulty, but those that combine structural, experiential and cultural qualities in a coherent way.

This finding is consistent with broader game studies arguments that digital games should be analysed as systems of rules, interactions and experiences rather than as isolated narrative or visual artefacts. The MDA framework proposed by Hunicke et al. (2004) is useful here because Ulduar’s strong performance can be interpreted as the alignment of mechanics, dynamics and aesthetics. Its encounter mechanics generate varied forms of player action; its progression structure creates dynamic group challenges; and its thematic coherence produces memorable aesthetic experience. Similarly, Sicart’s (2008) emphasis on mechanics as methods of interaction helps explain why encounter diversity and

design innovation are central to raid evaluation. A raid becomes memorable not only because of what it represents, but because of what it asks players to do collectively.

The upper-tier group consisting of Karazhan, Icecrown Citadel and The Nighthold further illustrates the multi-dimensional nature of raid quality. Karazhan performs strongly because of its distinctive atmosphere, structural originality and enduring community memory. Icecrown Citadel benefits from exceptional narrative integration and the iconic status of the Lich King as a final boss. The Nighthold stands out through its visual coherence, Suramar-centered narrative context and polished encounter structure. These raids do not excel for exactly the same reasons, yet they consistently appear near the top of the ranking. This supports the argument that raid quality is not a single construct but a composite of different design and cultural dimensions.

The results also show that the middle of the ranking is more methodologically sensitive than the top. Raids such as Black Temple, Throne of Thunder, Siege of Orgrimmar, Firelands and Battle of Dazar'alor display stronger rank variation across weighting schemes. This does not mean that these raids are weak; rather, their profiles are less uniformly balanced. For example, a raid may have high narrative importance but weaker accessibility, strong community memory but lower innovation, or high encounter complexity but less replay appeal. The observed mid-ranking instability is therefore substantively meaningful. It shows that different evaluative priorities can change how such raids are positioned.

The lower-ranked raids should also be interpreted carefully. Molten Core, Blackwing Lair and Sunwell Plateau have undeniable historical importance in the development of *World of Warcraft* raiding. Molten Core, in particular, is foundational because it introduced many players to large-scale raid progression. However, the model does not rank historical importance alone. It compares raids across ten criteria, including encounter diversity, atmosphere, replay appeal, accessibility burden and design innovation. From this perspective, some historically important raids score lower because their design profiles are narrower or less balanced by contemporary evaluative standards. This distinction is important because it prevents the results from being read as a simple rejection of early raid design. Instead, the findings show how historical importance and multi-criteria performance may diverge.

The high weight of C9 Accessibility Burden is also theoretically important. Since C9 is a cost criterion, its influence shows that excessive friction can negatively affect the overall evaluation of a raid. This has implications for how raid quality is understood. In player communities, difficulty is often treated as a marker of

prestige. However, the results suggest that difficulty and accessibility must be analytically separated. A raid can be challenging in a productive way, encouraging mastery, coordination and repeated learning; or it can impose burdens that reduce its broader experiential value. This distinction is consistent with MMORPG literature showing that online game experience involves achievement, social interaction and immersion simultaneously (Yee, 2006). If a raid's burden undermines social participation or replayability, its overall evaluation may decline even when it remains prestigious.

From a social and cultural perspective, the findings reinforce the importance of collective memory in MMORPGs. Raids are not consumed once and then forgotten; they are remembered, narrated, revisited and re-evaluated over time. This aligns with work on virtual worlds and collective memory, which suggests that persistent online environments can generate shared histories and durable cultural reference points (Papargyris & Poulymenakou, 2009). However, this study also shows that community memory should not be treated as the only indicator of quality. Nostalgia and historical visibility may elevate certain raids in player discourse, but a structured model can separate community memory from other dimensions such as innovation, replay value and accessibility.

Methodologically, the study demonstrates that MCDM methods can complement qualitative game analysis. Game studies has traditionally relied on textual analysis, ethnography, player studies, design theory and cultural interpretation. These approaches remain essential. However, MCDM adds a different analytical layer by making comparison explicit. CRITIC weighting identifies which criteria provide the strongest contrast in the dataset, while TOPSIS shows how each raid performs relative to an ideal and anti-ideal profile. The use of Entropy weighting, equal weighting, Spearman rank correlation and sensitivity analysis further improves transparency by showing whether the ranking depends on narrow methodological assumptions.

This methodological contribution is particularly relevant for the study of long-running digital games. Games such as *World of Warcraft* contain hundreds of comparable content units: raids, dungeons, zones, expansions, patches, classes, systems and progression models. Much of this content is evaluated informally in player communities, but formal academic evaluation remains limited. The framework proposed here can be adapted to other digital game contexts, such as dungeon design, expansion comparison, seasonal content evaluation, monetisation systems or platform-specific user experience. In this sense, the study contributes not only to *World of Warcraft* scholarship but also to the broader methodological toolkit of digital game studies.

The study also has practical implications for game design analysis. For designers, the results suggest that memorable raid content depends on balance rather than maximisation of a single dimension. A successful raid should combine meaningful mechanical variety, coherent narrative stakes, distinctive atmosphere, replay incentives and manageable accessibility. For scholars, the findings show that player-facing quality can be decomposed into analytically distinct but interrelated criteria. For player communities, the model offers a more transparent alternative to tier-list style ranking by showing how different assumptions about quality affect final rankings.

Nevertheless, the findings must be interpreted within the limits of the research design. First, several criteria were coded through structured desk evaluation rather than direct player survey data. Although this allows transparent and replicable modelling, future studies could strengthen the analysis by including multiple coders, expert panels or player-based rating data. Second, community memory is difficult to measure because it is shaped by nostalgia, player cohort, expansion timing and retrospective discourse. Third, the study evaluates twenty selected raids rather than the full universe of *World of Warcraft* raid content. Finally, the model is based on a specific criterion set. Alternative criteria, such as music, class balance, guild progression impact or world-first race visibility, could produce additional insights.

Despite these limitations, the study provides a clear answer to the research questions. MMORPG raid design can be operationalised as a multi-criteria evaluation problem by combining structural, experiential, narrative and community-memory criteria. When this framework is applied to selected *World of Warcraft* raids, Ulduar emerges as the most robustly superior alternative. The ranking is highly stable across alternative weighting schemes and sensitivity scenarios, especially at the top of the list. These findings support the usefulness of MCDM as a complementary approach for evaluating complex digital game content.

CONCLUSION

This study developed and applied a multi-criteria decision-making framework for evaluating selected *World of Warcraft* raids. The main objective was not to reproduce fan rankings or declare a universally definitive “best” raid, but to show how raid design can be analysed through explicit criteria, objective weighting, structured ranking and robustness testing. By combining CRITIC weighting with TOPSIS ranking, and by supplementing the baseline model with Entropy

weighting, equal weighting, Spearman rank correlation and $\pm 10/\pm 20$ sensitivity analysis, the study provides a transparent model for comparing raid content.

The results identify Ulduar as the strongest raid in the evaluated sample. Its first-place position remained unchanged across all weighting models and all sensitivity scenarios. This indicates that Ulduar's high ranking is not merely the result of one methodological assumption, but reflects a balanced performance across multiple criteria. Karazhan, Icecrown Citadel and The Nighthold also appear as consistently strong alternatives, forming a stable upper-tier group. In contrast, the middle section of the ranking shows greater sensitivity to weighting assumptions, suggesting that these raids have more uneven performance profiles.

The study contributes to the literature in three ways. First, it extends digital game studies by treating raid design as a structured evaluative problem rather than as an informal preference ranking. Second, it contributes methodologically by demonstrating how CRITIC–TOPSIS and related robustness techniques can be applied to digital game content. Third, it offers an empirical comparison of major *World of Warcraft* raids across criteria that capture design, narrative, atmosphere, replay value, accessibility and community memory.

The findings also show that raid quality is best understood as a balanced combination of design and cultural factors. A raid's value does not depend only on boss count, difficulty or historical importance. Instead, strong raid design emerges when mechanical diversity, narrative integration, aesthetic identity, replay appeal, innovation and accessibility are aligned. This is why historically important raids may not always rank highest, and why mechanically complex raids may be penalised if they impose excessive accessibility burden.

Future research can extend this study in several directions. First, the coding process can be strengthened through multiple coders and inter-coder reliability analysis. Second, player survey data or community-platform data can be incorporated to validate criteria such as community memory and replay appeal. Third, the sample can be expanded to include all *World of Warcraft* raids or comparative MMORPG cases such as *Final Fantasy XIV*, *Guild Wars 2* or *The Elder Scrolls Online*. Fourth, alternative MCDM methods such as EDAS, WASPAS, VIKOR or PROMETHEE can be used to test methodological robustness further. Finally, similar frameworks can be applied to other modular forms of digital game content, including dungeons, expansions, zones, game modes and monetisation systems.

Overall, the study shows that MCDM can provide a useful bridge between game studies, design analysis and structured evaluation. By making criteria and

weighting assumptions explicit, the proposed framework offers a transparent way to compare complex digital game content. In the case of *World of Warcraft* raids, the evidence points clearly to Ulduar as the most robustly high-performing raid in the evaluated sample, while also revealing how different raids achieve value through different combinations of design, memory and experience.

REFERENCES

- Aarseth, E. (2001). Computer game studies, year one. *Game Studies*, 1(1).
- Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, 22(7), 763–770.
- Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). “Alone together?” Exploring the social dynamics of massively multiplayer online games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 407–416.
- Golub, A. (2010). Being in the World (of Warcraft): Raiding, realism, and knowledge production in a massively multiplayer online game. *Anthropological Quarterly*, 83(1), 17–45.
- Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. *Proceedings of the AAAI Workshop on Challenges in Game AI*.
- Hwang, C. L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications: A state-of-the-art survey*. Springer.
- Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. MIT Press.
- Nardi, B. A. (2010). *My life as a night elf priest: An anthropological account of World of Warcraft*. University of Michigan Press.
- Papargyris, A., & Poulymenakou, A. (2009). The constitution of collective memory in virtual game worlds. *Journal of Virtual Worlds Research*, 1(3).
- Raudonat, K., & Marsden, N. (2018). Social actions in MMORPG raiding groups from the perspective of culture-inclusive action theory. *Proceedings of the 4th Conference on Gender & IT*, 101–111.

- Salen, K., & Zimmerman, E. (2004). Rules of play: Game design fundamentals. MIT Press.
- Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423.
- Sicart, M. (2008). Defining game mechanics. Game Studies, 8(2).
- Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72–101.
- Więckowski, J., Sałabun, W., & Wątróbski, J. (2023). Sensitivity analysis approaches in multi-criteria decision analysis: A systematic review. Applied Soft Computing, 148, 110915.
- Yee, N. (2006). Motivations for play in online games. CyberPsychology & Behavior, 9(6), 772–775.

Atf İçin: Bora, B. (2026). Digital Game Experience, Collective Memory and Raid Design: A Multi-Criteria Evaluation of World of Warcraft Raids, *Yeni Medya Elektronik Dergisi*, 10 (3), 457-488.