Original Research Article

Associations Between Video Game Features and IGN Review Scores

Ujaan Sanyal

American High School, 36300 Fremont Blvd, Fremont, CA 94536, United States

ABSTRACT

Video game review platforms such as IGN (Imagine Games Network) play a significant role in shaping consumer purchasing decisions and influencing the commercial success of games within the multibillion-dollar gaming industry. Despite the prevalent reliance on numerical review scores, existing research has largely ignored how specific game features—such as genre, Entertainment Software Rating Board (ESRB) age rating, or platform exclusivity— are statistically related to the scores these platforms assign. This paper examines how various features of video games are related to review scores on IGN, one of the most influential platforms in the gaming industry for developers and gamers seeking trusted reviews. Using a stratified random sample of 484 games weighted by inverse selection probability, one-way ANOVA and Tukey HSD tests with a Holm multiple-testing correction, and a weighted multivariable regression, this study evaluates which features of video games are associated with IGN scores. After employing these weighted methods along with omnibus tests, it was found that games developed in Japan scored higher than games developed in the United States, and M-rated games scored higher than E-rated games. Genre, platform of original release, platform exclusivity, and player mode showed no significant independent associations after adjustment, and the features together explained only a small proportion of score variance. Regarding release years, scores rose slightly across the interval. Understanding which game features are associated with higher IGN scores highlights patterns in review outcomes and helps developers and players identify which design characteristics are more likely to receive better feedback from IGN.

Keywords: ANOVA single-factor test; Tukey HSD test; IGN scores; Inverse-probability weighting; Weighted regression; Video games; ESRB; Genre

INTRODUCTION

Video games, dating back to 1958 when the first video game "Tennis for Two" was created by physicist William Higginbotham, have evolved from simple oscilloscopes to digital worlds that shape entertainment in the modern era (1). What began as a display to entice visitors has grown into a multibillion-dollar industry that continues to redefine entertainment and media as a whole. Since then, entertainment platforms such as IGN (originally known as the Imagine Games Network) act as a central hub for gaming news, community events, and most importantly, reviews for games made by IGN editors.

IGN began in 1996 as a collection of three console-specific websites. Over the years, IGN has merged its network of console-specific sites into one unified brand. Since the unification, IGN is now best known for its 10-point review system, where IGN scores of 0-1

Corresponding author: Ujaan Sanyal, E-mail: sanyalujaan@gmail.com.

Accepted August 21, 2026
https://doi.org/10.70251/HYJR2348.4413391349

represent the lowest range and scores of 9-10 represent the highest range (2). These score ranges now serve as a benchmark to gamers worldwide, shaping their decisions on which games they should buy and which are worth skipping out on. While the platform claims that reviews are subjective, it regularly rates games based on gameplay, replayability, and innovation. These factors influence the final score a game gets and provide insight into what the developers were focused on when creating the games (3).

Similar to other review systems, video game reviews have become an increasingly important factor in shaping consumer purchasing decisions, influencing the commercial success of a game (4). Review platforms such as IGN are widely recognized in the gaming industry for providing numerical ratings on games that players use to determine whether or not they should buy a game. Despite the large influence these review scores have over customer purchases, research remains limited on the impacts certain game characteristics and features have on overall IGN scores. Research regarding gaming journalism has found that discussions involving gaming events and culture such as Gamergate have had a significant impact in shaping public opinion and perception around the credibility of gaming media. These perceptions have contributed to the rise of ongoing debates regarding the legitimacy of gaming media, including review sites like IGN (5). Additional studies have also closely examined the structure and priorities of game journalism and compared their characteristics to traditional journalism. For instance, research comparing gaming journalism to traditional journalism has found differences in content priority and evaluation criteria, suggesting that the methods video game journals use to evaluate games may be influenced beyond a game's technical quality (6).

Although prior studies have investigated the effects of gaming journalism and its influence on public opinion when compared to traditional journalism, limitations on the understanding of the statistical factors associated with review scores do exist. Existing literature has focused only on journalistic practices and public perceptions, rather than on the statistical relationships between game features and review outcomes. This gap motivates the current study, which examines the specific game features that may be associated with a game's IGN score through a stratified random sample, including ESRB rating, genre, platform of original release, player mode, platform exclusivity, and country of development. These features are analyzed through statistical tests to identify associations in the results. It is hypothesized that games that were not platform exclusives will receive higher average IGN review scores than platform exclusives, because of the possible market pressures multi-platform titles face to cater to a wider audience across different consoles, ultimately leading to a stronger reception from reviewers.

METHODS AND MATERIALS

Data Collection

This study involved a Kaggle dataset containing 12,589 unique video game titles from the years 1970-2016 and their corresponding IGN score (7). Features in this dataset include title, IGN score, platforms launched, genre, and the release year. Unique titles were identified by the exact name, so that the same title released on multiple platforms was only counted once, while distinct names (including special editions and limited editions) were counted distinctly. A stratified random sample was performed of 484 games from the Kaggle dataset evenly across each IGN score range: 34 games from (0-1], 50 games from (1-2], 50 games from (2-3], 50 games from (3-4], 50 games from (4-5], 50 games from (5-6], 50 games from (6-7], 50 games from (7-8], 50 games from (8-9], and 50 games from (9-10]. Note that the disparity in games from the IGN score range (0-1] was due to the lack of unique titles in the Kaggle dataset for that category. For each score bracket, the sampling frame consisted of all unique titles with at least one platform-specific IGN review in that bracket. A title was only counted once within a given bracket, but it could appear in more than one bracket if the game had varying reviews (fell into different score brackets) for different platforms. By this definition, the (0-1] bracket contained exactly 34 unique titles as mentioned before. However, across brackets, a title only contributed a single score and only appeared once, so observations were independent. Simultaneous multi-platform releases were classified as "Multiple" in the platform of original release column while for exclusives, ports, remasters, and other rereleases (that shared the exact same title as the original), the platform of original release referred to the earliest platform on which the underlying game was released. After each game was randomly sampled, the data was then extracted to a table containing all the features in the original Kaggle dataset, as well as newly added features such as platform of original release, country of development, Entertainment Software Rating Board (ESRB) rating, exclusivity on release, and player

mode. Information regarding these excess features was found either on IGN's website, MobyGames, or on IGDB's (Internet Game Database) video game database, containing information for over 400,000 titles (8-10).

Within each score bracket, a random number was generated for every eligible title using the Google Sheet's random-number function. Then, the rows were sorted by that value, where the first 50 titles were selected, except in the (0-1] bracket, where all 34 titles were included. A random seed was not recorded, so the exact draw cannot be reproduced, although the full bracket structure of the sampled games is fully documented in Table 1. Regarding how information was collected for features outside of the original Kaggle dataset, IGN was used first, then MobyGames, and then IGDB. Manual coding for features were independently checked (not by the original coder) and confirmed for accuracy. Although the source dataset spans from 1970-2016, nearly all sampled games were released between 1995 and 2015. For the seven games that were released outside that interval, they were only excluded from Figure 1.

Operational Definitions

Two coded features warrant explicit definition. The country of development was defined as the country in which the game's primary or lead development studio was based. For multinational studios or co-developed titles, the country of the studio identified by IGN, MobyGames, or IGDB was used. Platform exclusivity was coded "Yes" if the game was available on only one platform upon release, which includes titles whose exclusivity was temporary, and coded "No" otherwise.

Missing Data and Category Thresholds

Player mode could only be clearly determined for 100 of the 484 games (20.7%); of those 100, 2 games were strictly multiplayer-only games and were excluded due to the small size of the category, meaning only 98 games (52 single-player, 46 combined single-player and multiplayer) were analyzed. Optional online functions, such as leaderboards, downloadable content, and account connectivity, were not treated as multiplayer unless the game included a playable multiplayer mode where players can interact with other players. Because games with missing player mode information (the majority of games) may differ systematically from those with it, the player mode analysis is interpreted with particular caution and speculation. In the platform comparison, the 134 games that launched simultaneously on multiple platforms were excluded, since no original platform could be assigned to them. Regarding the genre comparison and platform of original release comparison, only the four most popular genres and platforms were included, as the prespecified minimum category threshold was 25 games; smaller

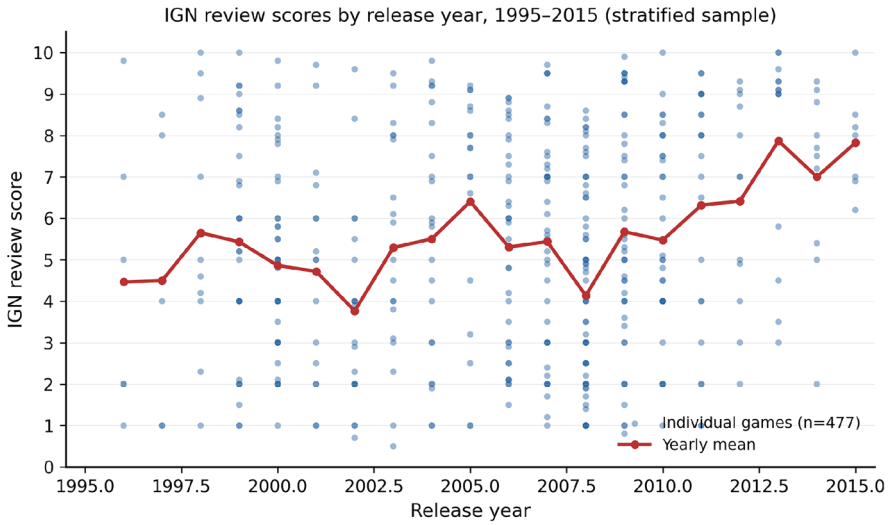

Figure 1. *The scatterplot displays individual game review scores (y-axis) plotted against their release year (x-axis) for all titles included in the stratified random sample except 7 games that fell outside this range (n = 477). Each blue point represents a single game's IGN score. The distribution of points illustrates overall scoring patterns across two decades (1995-2015). This is highlighted by the red line, which shows the yearly mean. Yearly means are unweighted within the outcome-stratified sample.*

categories were excluded rather than tested individually. Furthermore, 11 games were not analyzed in the ESRB comparison as they were unrated. The rest of the features, country of development and platform exclusivity, were analyzed for all 484 values. Due to missing or selectively available information, especially for player mode, genre, and platform of original release, conclusions drawn from those tests should be interpreted with caution and regarded as tentative rather than as firm associations.

Tests Performed

Using these features, two statistical analysis tests were performed to analyze the relationship between each game feature and its corresponding IGN review score. The first test, the ANOVA single-factor test, was used to determine whether a statistically significant difference can be found among two or more independent groups by comparing their mean values (11). ANOVA tested the null hypothesis that all group means are equal by comparing the variance between a group to the variance within that group. A statistically significant result ($p < 0.05$) indicated that at least one group mean was significantly different from one of the others (12). The second test, the Tukey Honestly Significant Difference (HSD) post-hoc test, was used after a significant difference was found in the ANOVA test to identify exactly which pair of groups differed significantly from each other (13-14). These two tests form the exploratory core of the analysis; they are complemented with assumption diagnostics (Welch's ANOVA, Levene's test, Shapiro-Wilk test), the Holm multiple-testing correction, and least-squares weighted multivariable regression. Together, these tests helped identify which features of a video game were associated with its IGN score and identified which groups within those features differed significantly. To ensure an adequate sample size, categories were only used if they had a sufficient number of values (at least 25). For example, while performing the tests on the genre category, only the four most popular genres (action, racing, shooter, and sports) were tested.

Sampling Weights

Because an equal number of games were chosen from score brackets (50 games from each score bracket) of different sizes, the unweighted sample over-represented low-scoring games (games from the (0-1] bracket) relative to the full dataset. If left unchanged, this unequal sampling distribution may have led to distortions in category means, variances, effect estimates, and significance tests. To account for this, each game was assigned an inverse-probability weight equal to the number of unique titles in the Kaggle dataset over the ones that were sampled (usually 50). For instance, the (0-1] dataset only had 34 unique titles in the dataset, and since all 34 were sampled, its inverse-probability weight was 1.00.

Primary Analysis

The primary analysis was a weighted least-squares regression of IGN score on genre (Action, Racing, Shooter, Sports, Other), country of development (United States, Japan, Other), ESRB rating (E for Everyone, T for Teen, M for Mature), platform exclusivity, and release year. Heteroscedasticity-robust (HC3) standard errors were used. Player mode was excluded for this analysis as it lacked information for most games. Platform of original release was similarly excluded because it showed no significant association in any test. The one-way ANOVA tests and the Tukey HSD tests serve as more of an exploratory analysis, compared to the weighted regression, which serves as the primary analysis.

Assumption Checks, Multiple Testing, and Software

Homogeneity of variance was assessed with Levene's tests (median-centered). Furthermore, Welch's ANOVA, which does not assume that all groups have equal variances, was also reported for each factor to check for robustness. Residual normality was evaluated using the Shapiro-Wilk test. Because six separate omnibus tests were conducted, Holm-adjusted p-values were reported alongside the unadjusted p-values to control the family-wise error rate (15). For these omnibus tests, statistical significance was defined as $p < 0.05$. All analyses were computed in Python 3.12 (Google Colab) using scipy 1.16.3 (16) and statsmodels 0.14.6 (17).

RESULTS

Sampling Weights and Statistic Analysis

The weights for each score bracket (number of total values divided by the number of sampled values) ranged from 1.00 in the (0-1] bracket, which was fully represented by the sampled data, to 75.82 in the (7-8] bracket (Table 1). To validate the weighting, the weighted sample mean score was 6.86, matching the mean score over the bracket frames (6.86), when rounded off to the nearest hundredth, whereas the unweighted sample mean was 5.40.

Group sizes, means, and standard deviations were computed for every feature analyzed (Table 2). However, only the most popular values in the feature were analyzed

for genre and platform of original release. Across all 484 sampled games, the mean IGN score was 5.40 and the standard deviation (SD) was 2.70.

Omnibus Tests and Assumption Checks

Significant overall differences were found for genre, $F(3, 283) = 2.95$, $p = 0.033$; country of development, $F(2, 481) = 5.45$, $p = 0.005$; ESRB rating, $F(2, 470) = 4.25$, $p = 0.015$; and platform exclusivity $F(1, 482) = 7.50$, $p = 0.006$. No significant differences were found for platform of original release, $F(3, 156) = 2.06$, $p = 0.108$, or player mode, $F(1, 96) = 3.33$, $p = 0.071$ (Table 3). Effect sizes remained small throughout the various tests, with eta-squared ranging from 0.015 to 0.038. Levene's test indicated homogeneous variances for all six factors (all $p > 0.44$), meaning that the assumption of equal variances required by the one-way ANOVA was fulfilled, and Welch's ANOVA resulted in the same conclusion. Shapiro-Wilk tests indicated departures from residual normality for all six factors (W between 0.946 and 0.967; $p \leq 0.006$ for all factors; Table 3). However, given the group sizes and HC3 robust standard errors in the primary regression, these results were not considered enough to invalidate the main conclusions.

Table 1. *Stratified sample structure and inverse-probability weights. N counts unique titles (matched by exact name) with at least one review score in the bracket. Titles with scores in more than one bracket belong to each applicable frame, meaning that the N columns total 14,414 instead of 12,589 unique titles in the dataset. Weighted sample mean = frame mean = 6.86; unweighted sample mean = 5.40.*

Score bracket	Unique titles in bracket (N)	Titles sampled (n)	Weight (N/n)
(0, 1]	34	34	1.00
(1, 2]	167	50	3.34
(2, 3]	431	50	8.62
(3, 4]	741	50	14.82
(4, 5]	1,238	50	24.76
(5, 6]	1,757	50	35.14
(6, 7]	2,764	50	55.28
(7, 8]	3,791	50	75.82
(8, 9]	2,856	50	57.12
(9, 10]	635	50	12.70
Total	14,414	484	—

Multiple-Testing Correction and Post-Hoc Tukey HSD Comparisons

Since six omnibus tests were conducted, Holm adjusted p-values were computed for them. After computation, it was seen that country of development (adjusted $p = 0.027$) and platform exclusivity (adjusted $p = 0.032$) remained significant, while ESRB rating (adjusted $p = 0.060$) and genre (adjusted $p = 0.099$) did not (Table 4).

Tukey HSD comparisons with mean differences and 95% confidence intervals were computed for factors with

Table 2. *Descriptive statistics for every group analyzed. ESRB analysis excludes 11 of the 484 games as they were unrated. Both genre and platform release are restricted to the 4 most common categories. The platform of original release statistics merged Windows and PC (n = 160), excluding games that launched simultaneously on both platforms. Across all 484 games, the mean score was 5.40 and the Standard Deviation (SD) = 2.70.*

Feature	Group	n	Mean	SD
Genre	Action	123	5.44	2.68
	Racing	46	4.21	2.60
	Shooter	46	5.55	2.72
	Sports	72	4.81	2.94
Country of development	United States	190	5.04	2.65
	Japan	126	6.04	2.62
	Other Countries	168	5.31	2.73
ESRB rating	E	278	5.10	2.69
	T	123	5.68	2.57
	M	72	6.01	2.87
Platform of original release	PC	55	5.32	2.56
	DS	29	5.08	2.15
	PS1	30	4.26	2.40
	PS2	46	4.28	2.66
Platform exclusivity	Yes	243	5.06	2.67
	No	241	5.73	2.69
Player mode	combined single-player and multiplayer	46	5.90	2.76
	Single-player	52	4.88	2.75

three or more groups, coupled with the equivalent two-group comparisons of platform exclusivity and player mode (Table 5). From the tests, two pairwise differences reached significance: action games scored 1.24 points higher than racing games (95% CI [0.01, 2.46], adjusted p = 0.046), and games developed in Japan scored 1.00 point higher than games developed in the United States (95% CI [0.28, 1.72], adjusted p = 0.003), with group means of 6.04 and 5.04 points respectively (Table 2). Regarding ESRB, M-rated games scored 0.91 points higher than E-rated games (95% CI [0.08, 1.75], adjusted p = 0.029), with group means of 6.01 and 5.10 points. No pairwise platform comparison was found significant. For platform exclusivity, non-exclusive games scored 0.67 points higher than exclusive games (means of 5.73 versus 5.06; 95% CI [0.19, 1.14], p = 0.006), and for the select-few games ranked by player mode, the difference of 1.02 points was not significant (95% CI [-0.09, 2.13], p = 0.071).

Weighted Multivariable Regression

The primary weighted least-squares regression only included 473 games that fit the E, M, or T ESRB rating, since 11 out of the 484 games were unrated (Table 6). In the weighted multivariable regression, games developed in Japan scored an estimated 0.60 points higher than games developed in the United States (b = 0.60, 95% CI [0.21, 0.98], p = 0.003), and M-rated games scored 0.67 points higher than E-rated games (b = 0.67, 95% CI [0.21, 1.13], p = 0.005), after adjustment for the other variables in the model. Furthermore, scores increased by an estimated 0.04 points every release year (95% CI [0.005, 0.078], p = 0.027). No genre category differed significantly from action games, and platform exclusivity was not significantly associated with score after adjustment (b = −0.20, 95% CI [−0.54, 0.13], p = 0.24). The model explained a small proportion (approximately 6%) of the variation in IGN scores ($R^2 = 0.06$).

Table 3. *This table features combined tests and assumption diagnostics. Levene's test (median-centered) was non-significant for every factor, indicating homogeneous variances. Welch's ANOVA (reported here as its p-value) further corroborated that result in every case. The Shapiro-Wilk test on the ANOVA residuals was significant for every factor, indicating a departure from normality. Eta-squared values indicate small effect sizes throughout. Holm-adjusted p-values are excluded here and instead recorded in Table 4.*

Feature	n	One-way ANOVA	p	η^2	Levene p	Welch's ANOVA	Welch p	Shapiro–Wilk W (p)
Genre	287	F(3, 283) = 2.95	0.033	0.030	0.526	F(3, 119.5) = 3.10	0.029	0.946 (p < 0.001)
Country of development	484	F(2, 481) = 5.45	0.0046	0.022	0.813	F(2, 301.3) = 5.61	0.0041	0.957 (p < 0.001)
ESRB rating	473	F(2, 470) = 4.25	0.015	0.018	0.446	F(2, 173) = 4.06	0.019	0.958 (p < 0.001)
Platform of original release	160	F(3, 156) = 2.06	0.108	0.038	0.593	F(3, 77.9) = 1.98	0.123	0.967 (p < 0.001)
Platform exclusivity	484	F(1, 482) = 7.50	0.0064	0.015	0.602	F(1, 481.9) = 7.50	0.0064	0.959 (p < 0.001)
Player mode	98	F(1, 96) = 3.33	0.071	0.034	0.844	F(1, 94.5) = 3.33	0.071	0.962 (p = 0.006)

Table 4. *Holm correction applied across all six omnibus ANOVAs. Genre and ESRB rating were found not significant after the Holm correction (though they were initially significant from the omnibus tests) alongside player mode and platform of original release. Only country of development and platform exclusivity remained significant after Holm correction.*

Feature	Raw p	Holm-adjusted p	Significant at α = 0.05
Country of development	0.0046	0.027	Yes
Platform exclusivity	0.0064	0.032	Yes
ESRB rating	0.015	0.060	No
Genre	0.033	0.099	No
Player mode	0.071	0.142	No
Platform of original release	0.108	0.142	No

Table 5. *This table contains pairwise comparisons across categories in a feature. For features with three or more categories, Tukey HSD mean differences, 95% confidence intervals, and Tukey adjusted p-values are reported. For platform exclusivity and player mode, which only have two groups, the mean difference, its 95% confidence interval, and the ANOVA p-value are reported, and no post-hoc test is required. For every comparison, the mean difference is reported as the first group minus the second, so that a positive value indicates that the first group scored higher. Confidence intervals follow the same direction.*

Feature	Comparison	Mean diff.	95% CI	Adjusted p	Significant
Genre	Action vs Racing	1.24	[0.01, 2.46]	0.046	Yes
	Action vs Shooter	-0.11	[-1.33, 1.12]	0.996	No
	Action vs Sports	0.63	[-0.42, 1.68]	0.410	No
	Racing vs Shooter	-1.35	[-2.82, 0.13]	0.088	No
	Racing vs Sports	-0.61	[-1.95, 0.73]	0.644	No
	Shooter vs Sports	0.74	[-0.60, 2.08]	0.484	No
Country of development	Japan vs United States	1.00	[0.28, 1.72]	0.0034	Yes
	Japan vs Other Countries	0.73	[-0.01, 1.47]	0.053	No
	Other Countries vs United States	0.27	[-0.40, 0.93]	0.609	No
ESRB rating	M vs E	0.91	[0.08, 1.75]	0.029	Yes
	T vs E	0.58	[-0.11, 1.26]	0.118	No
	M vs T	0.34	[-0.60, 1.27]	0.677	No
Platform of original release	DS vs PC	-0.24	[-1.72, 1.24]	0.975	No
	DS vs PS1	0.82	[-0.86, 2.50]	0.587	No
	DS vs PS2	0.80	[-0.73, 2.33]	0.530	No
	PC vs PS1	1.06	[-0.41, 2.53]	0.242	No
	PC vs PS2	1.04	[-0.25, 2.33]	0.161	No
	PS1 vs PS2	-0.02	[-1.54, 1.50]	1.000	No
Platform exclusivity	No vs Yes	0.67	[0.19, 1.14]	0.0064	Yes
Player mode	Combined single-player and multiplayer vs Single-player	1.02	[-0.09, 2.13]	0.071	No

Table 6. *Weighted least-squares regression of IGN score on game features and release year, excluding 11 unrated ESRB games. (n = 473; inverse-probability weights by score bracket; HC3 heteroscedasticity-robust standard errors; $R^2 = 0.06$). Release year is centered on the sample mean (2006.5) Player mode was excluded from this model due to missing data (player mode was only known for 98 games) and instead was analyzed by one-way ANOVA in Tables 3-5 above.*

Term	b	Robust SE	95% CI	p
Intercept	6.29	0.25	[5.81, 6.77]	< 0.001
Genre: Racing (vs Action)	0.04	0.34	[-0.64, 0.71]	0.917
Genre: Shooter (vs Action)	0.16	0.33	[-0.48, 0.80]	0.624
Genre: Sports (vs Action)	0.41	0.32	[-0.22, 1.05]	0.199
Genre: Other (vs Action)	0.28	0.20	[-0.12, 0.67]	0.175

Continued Table 6. *Weighted least-squares regression of IGN score on game features and release year, excluding 11 unrated ESRB games. (n = 473; inverse-probability weights by score bracket; HC3 heteroscedasticity-robust standard errors; R^2 = 0.06). Release year is centered on the sample mean (2006.5) Player mode was excluded from this model due to missing data (player mode was only known for 98 games) and instead was analyzed by one-way ANOVA in Tables 3-5 above.*

Term	b	Robust SE	95% CI	p
Country: Japan (vs United States)	0.60	0.20	[0.21, 0.98]	0.0026
Country: Other (vs United States)	0.25	0.20	[-0.14, 0.64]	0.214
ESRB: T (vs E)	0.28	0.19	[-0.09, 0.65]	0.138
ESRB: M (vs E)	0.67	0.24	[0.21, 1.13]	0.0046
Exclusive at release: Yes (vs No)	-0.20	0.17	[-0.54, 0.13]	0.236
Release year (per year, centered)	0.04	0.02	[0.01, 0.08]	0.027

DISCUSSION

General Trend

Figure 1 shows the progression of IGN scores throughout the years (1995-2015) based on the sample games in the dataset. In this figure, games from the 2010-2015 range had a higher average IGN score (note that 7 out of the 484 games were not plotted as they were reviewed outside of the time range). This may have been due to the shift in quality of video games during this time period, such as higher-definition graphics and increased console accessibility, which appealed to a wider audience. Furthermore, this increase may also have been a result of commercial globalization, which expanded the reach of video games into broader international markets, possibly elevating visibility among reviewers. The regression model reflects this trend as a small positive association between release year and IGN score.

Genre Comparisons

This ANOVA test was taken as a single factor between the four most populated values (action, racing, shooter, sports) in the 'genre' column. The omnibus test provided evidence of a difference among the four categories. Follow-up Tukey HSD comparisons demonstrated that action games scored significantly higher than racing games, whereas no other pairwise genre comparisons reached statistical significance.

These statistical analyses revealed patterns in the scores IGN reviewers assigned across game features. On average, action games were rated higher than racing games, and this was the only genre pair to differ significantly in the post-hoc comparisons. A possible interpretation, which the study cannot test directly, is that action games offer a much deeper narrative, in terms of story and in-game experiences, which may have resonated stronger with IGN reviewers than the repetitive races involved in most racing games. However, this difference remains tentative as the genre omnibus test is no longer statistically significant after Holm adjustment across the six factors, and no genre category differed significantly from action games once the other features and release year were adjusted for. The analysis was restricted to four genre categories due to sample-size constraints, meaning that future research including additional genres or a larger dataset of existing genres may reveal further significant differences.

Country Comparisons

An ANOVA test was then performed on the country where games were developed, using the following categories: US-made games, Japan-made games, and games made in any other country. Evidence for a statistical difference could be found, meaning that Japan-made games had a significant difference in score when compared to the US-made games. The Tukey test comparing the country of origin of the games found a statistical difference between games developed in the United States and those developed in Japan, while comparisons involving the "Other Countries" category were not statistically significant.

The ANOVA analysis of the countries where games were developed revealed that Japan-made games outscored US-made games on average. Furthermore, this significance still stands after Holm adjustment, Welch's test, and the weighted adjusted model. The

following explanations are possible interpretations rather than confirmed findings. This gap may reflect the innovative and artistic ideas represented in Japanese game development, which appeared to align strongly with IGN review preferences. Brand recognition may also have played a role, as well-established Japanese franchises like Nintendo were known for making games with iconic characters and creative worlds. For instance, popular characters like Mario and Bowser may have had a stronger connection with IGN reviewers due to their long-standing presence in the media, potentially leading to more favorable evaluations of Japanese-developed titles overall.

ESRB Rating Comparisons

A one-way ANOVA demonstrated a significant difference in IGN scores among ESRB rating categories. Subsequent Tukey HSD analysis indicated that the only significant pairwise comparison was between M-rated games and E-rated games, with M-rated games scoring higher on average. A possible explanation for this may be that E-rated games are aimed at more casual audiences and include a large number of lower-budget games, which IGN reviewers may have evaluated less favorably compared to the more narratively ambitious titles often attributed to M-rated games. It should be noted, however, that the ESRB omnibus test did not remain significant after the Holm correction, meaning this pattern should remain tentative and should not go beyond general speculation.

Platform Of Original Release Comparisons

After merging the duplicate PC and Windows categories, the platform analysis found no significant overall difference and no significant pairwise differences. Consequently, the results do not provide strong enough evidence that the platform of original release is associated with IGN review scores. A possible interpretation, which this study cannot test directly, is that reviewer evaluations track qualities of the individual game rather than the hardware on which it was introduced, so that no platform stands out once the same titles are compared on a uniform metric.

Platform Exclusivity on Release Comparisons

Non-exclusive games received higher average IGN scores than platform-exclusive titles. A plausible explanation for this partial support is that games released on multiple platforms at launch allow for higher accessibility to a wider audience, who may not have the specific platform required for a platform-exclusive title. However, this explanation is speculative as this study did not analyze factors like audience size, production budget, and development quality. Furthermore, although the unadjusted difference remained significant after the Holm correction, platform exclusivity was not clearly associated with IGN scores after adjustments for genre, country of development, release year, and ESRB rating. The apparent advantage of non-exclusive games may therefore be partly due to these other characteristics. Overall, the hypothesis received support in the unadjusted difference but not in the adjusted regression.

Player Mode Comparisons

A one-way ANOVA was conducted to evaluate differences between 46 combined single-player and multiplayer games (containing both single-player and multiplayer modes) and 52 single-player-only games. This size discrepancy was due to a lack of information for most of the remaining games on notable video game databases. The analysis did not reveal a statistically significant difference between the two groups.

In this comparison, which included only a subset of games with a known player mode, games that featured both multiplayer and single-player aspects received higher average IGN scores than games that were only single-player, although this difference was not statistically significant. A plausible explanation for this minor difference may be that combined single-player and multiplayer games offer greater replayability and offer more options to different kinds of players. However, because the difference was not statistically significant, the 95% confidence interval for this difference includes zero, and player mode could only be determined for 98 of the 484 games, these explanations remain speculative and no conclusions should be drawn from the data.

CONCLUSION

This study examined how features of a video game, such as genre, platform of original release, platform exclusivity, ESRB rating, player mode, and country, relate to IGN reviewer scores. Using stratified sampling with inverse-probability weights and statistical testing, the clearest association was found for the country of development, with Japan-developed games scoring higher than US-developed games in both unadjusted and adjusted analyses. Similarly for ESRB rating, M-rated games also scored higher than the E-rated games in the

adjusted model. However, genre and ESRB differences did not survive the multiple-testing (Holm) correction, and no differences were found for platform of original release or player mode. Finally, the original hypothesis that non-exclusive games would receive higher IGN scores than exclusives was only partially supported, as the unadjusted difference was significant but it did not remain so after adjustment for the other features and release year. While reviews remained inherently subjective, these distinctions highlighted the potential patterns IGN reviewers favored when rating games.

Several limitations should be explored and acknowledged beyond the small sample size. The outcome-stratified sampling design required inverse-probability weighting, which corrects estimates in expectation, but is not as precise as sampling based on an equal simple random size. Furthermore, player-mode information was missing for most games, several analyses were restricted to the most common categories, and the manually coded variables may contain misclassification even after independent checking. Six exploratory tests increased the risk of false positives, which the Holm correction addresses, and unmeasured characteristics like franchise status, publisher, and production budget may serve as possible confounding factors for the remaining associations. Finally, all scores come from a single review platform over a historical period ending in 2016, so the observed patterns may not generalize to other outlets or to recent games.

Other modifications for future research could include analysis for a wider collection of features, such as length of the game, art style/animation, and whether or not the game is part of a continuing franchise. Stemming from this previous idea, it may be interesting to compare games from a well-established franchise or part of a series of games to standalone games that the audience has no prior background with. Future research could also draw a random representative sample of games, multivariable analysis, temporal adjustment, and comparison across a variety of review outlets. Overall, these findings highlight the importance of understanding how specific design and production choices are associated with the critical reception reviewers assign to games while also offering a foundation for deeper investigation into the systematic patterns that shape video game evaluation.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Won Park for guidance and mentorship throughout this research.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

RFERENCES

1. October 1958: Physicist Invents First Video Game. APS News. American Physical Society. Available from: https://www.aps.org/apsnews/2008/09/first-video-game (accessed on 2026-01-27).
2. Geddes R. Origins: The History of IGN. IGN. Available from: https://www.ign.com/articles/2008/01/11/origins-the-history-of-ign (accessed on 2026-01-27).
3. IGN Entertainment. IGN Review Practices. Available from: https://corp.ign.com/review-practices (accessed on 2026-01-27).
4. Alexander L. Study: Review Scores Directly Impact Consumer Purchasing, Opinion. Game Developer. 2010 Jul 7. Available from: https://www.gamedeveloper.com/game-platforms/study-review-scores-directly-impact-consumer-purchasing-opinion (accessed on 2026-01-27).
5. Perreault G, Vos T. Metajournalistic discourse on the rise of gaming journalism. *New Media Soc*. 2020; 22 (1): 159–176. https://doi.org/10.1177/1461444819858695
6. Almqvist J, Bartolic H. Games journalism — a deserved bad reputation? Jönköping University. Available from: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1889520 (accessed on 2026-01-27).
7. Kapturov A. IGN Games From Best to Worst. Kaggle. Available from: https://www.kaggle.com/datasets/kapturovalexander/ign-games-from-best-to-worst (accessed on 2025-01-03).
8. IGN Entertainment. All Reviews. IGN. Available from: https://www.ign.com/reviews (accessed on 2026-01-27).
9. MobyGames. Video Games Database. Available from: https://www.mobygames.com (accessed on 2026-01-27).
10. Internet Game Database. Advanced Search. IGDB. Available from: https://www.igdb.com/advanced_search (accessed on 2026-01-27).
11. Kent State University Libraries. SPSS Tutorials: One-Way ANOVA. Available from: https://libguides.library.kent.edu/spss/onewayanova (accessed on 2026-01-27).
12. Frost J. Null Hypothesis for ANOVA. Statology. Available from: https://www.statology.org/null-hypothesis-for-anova (accessed on 2026-01-27).
13. Scagnoli A. Post Hoc Analysis: Tukey HSD Test. RPubs. Available from: https://rpubs.com/aaronsc32/post-hoc-analysis-tukey (accessed on 2026-01-27).
14. Vasavada N. ANOVA With Post-hoc Tukey HSD Test Calculator. Astatsa.com. Available from: https://astatsa.com/OneWay_Anova_with_TukeyHSD (accessed on

2026-01-27).
15. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics [Internet]. 1979 Jan 1; 6: 65–70. Available from: https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf (accessed on 2026-01-27).
16. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods [Internet]. 2020 Feb 3; 17 (3): 261–72. Available from: https://doi.org/10.1038/s41592-019-0686-2 (accessed on 2026-01-27).
17. Seabold S, Perktold J. statsmodels: econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference. 2010; 92–96. https://doi.org/10.25080/Majora-92bf1922-011