



Metacritic: Quantify Me

By Mike W. Zuurman

January 2014

XeNTaX Foundation

<http://www.xentax.com>

v1.0

Abstract

Background. The *Metacritic* website (<http://www.metacritic.com>) is considered a reference point where collected scores from critics and users can be found regarding different types of media released to the public. Such, that these *metascores* may even influence consumer decisions to buy those media releases. It has been suggested that *Metacritic* converts each review into a percentage the site decides for itself, that it manually assesses the tone of reviews that have no explicit score and then assigns a quantitative score and finally that it refuses to reveal what weights are applied to which publications. Assuming the above is correct; this would make the data shown at *Metacritic* highly biased and possibly far from the truth. The objective of my current analysis is to visualize the data regarding *Computer Games*, compare critic scores with user scores, in time and cross-sectional, look for patterns and then try to explain some of those patterns. I have one primary hypothesis, and that states that both critic metascores and user metascores differ significantly, based on the idea that users rate games differently than the “professional reviewers”, especially considering the scoring method of *Metacritic* is suggestive of statistical flaw. **Methods.** Between 25th of August 2013 and the 10th of September 2013 a custom made tool called Slurp! extracted the following data from the Metacritic website (<http://www.metacritic.com>). For each game it collected the Title, Publisher, Developer, Release Date, Platform, Critic Metascore, Number of Critics, User Metascore, Number of Users, Individual Critic Reviews (including Critic Name and links to the original reviews if present), and Individual User Reviews (including User Name and review text of user). This data was then fed to IBM Statistics 20.0 for further analysis. **Results.** The total number of games listed at *Metacritic* was 136.496 at the time of database completion. Game review scores from 14 different platforms were collected. Both the *Critic score* (CS) and *User score* (US) data was *skewed* and not distributed normally according to the Kolmogorov-Smirnov Test for all platforms. For each platform there is a significant difference between the median US and median CS, except the 3DS, Playstation Vita and Xbox 360. Overall, users rated 54% of games higher than the critics, rated 41% of games less than the critics, and showed agreement with critics in 5% of the cases. In all, User and Critic scores differ significantly from each other (standardized test statistic: 14.623, $P < 0.0001$). In a univariate linear model, the year of release is significantly associated with User score ($P < 0.0001$), with a B of -1.036 and R^2 of 6.3%. Thus, US for games get lower with advancing years. The median CS by platform also differ. For instance, the “population” of critics at Metacritic rate Wii games the lowest of all, while in contrast, iOS games seem to be rated among the highest of all. Of note: 112243 iOS games are listed at *Metacritic*, but only 2% of those have any CS – 98% of the iOS games are listed without any score. US for Playstation 1 games are rated the highest versus all other platforms, while again Wii games are rated the lowest by users. Also high are Gameboy Advance and Playstation 2 games, while on the lower side we find Xbox 360, Playstation 3 and 3DS. On the whole, the data suggests games on older platforms are rated higher by users than games for the new platforms. The number of critic ratings and user ratings are also different between platforms as Xbox 360 and PC games have the most ratings, while Wii U and Playstation 1 games are among the least rated at

Metacritic. The number of games listed by year have been examined as well. Games from 1996 number 299 and while there is a steady increase in the number of games listed at *Metacritic* by release year, 2009 shows a dramatic increase to 18122 (from 3842 in 2008). The number of reviews per game (both Critic and User reviews) were plotted against release year. Overall, there is a steep dramatic drop in median number of Critic reviews from 2006 onwards, mainly caused by a drop in median number of Critic reviews for Xbox and Playstation 2 games. When looking at the top 20 games in terms of CS, it is noteworthy that these games were rated much higher by *Critics* as opposed to Users pointing to overrating of games by Critics. These data suggest that games like Company of Heroes 2, Diablo III, COD:MW2&3, LittleBigPlanet, SimCity and Mass Effect 3 have been seriously highly overrated, as far as gamers (=Users at *Metacritic*) are concerned. Indeed, there is more overrating by *Critics* than underrating when looking at the extremes. For example, in terms of publishers, Electronic Arts, Activision, Blizzard Entertainment and Rockstar are among those where critic scores deviate most from users, being higher than the US. 285341 reviews from 366 unique critics are listed at *Metacritic*. Of these, the top 5 listed critics are IGN, GameSpot, GameZone, Cheat Code Central and Game Informer, together contributing 13% of all reviews. Xbox games have approximately 25 reviews (by Critics) per game at *Metacritic*, while PC games and iOS games have the lowest ratio (5.5 and 0.1 review per game respectively). *Metacritic* is heavily biased to console games. **Conclusion.** The data at *Metacritic* leaves much to be desired and seems to be heavily biased. When critic scores do not comply with user scores in the majority of cases, which has been shown in this paper, and the selection of critics is seemingly under no formal control, the validity and accuracy of the data is low. Caution is necessary when using *Metacritic* information to guide your own decision to buy or not buy a game. Until there is more transparency on how this process takes place at *Metacritic*, more transparency on the flow of funding from which sources, and the observed biases are removed, the database is of limited use for the end-users.

1. Introduction

As an introduction to the current paper, let's first read what Wikipedia has to say about Metacritic dating today, the 16th of September 2013.

"Metacritic is a website that aggregates reviews of music albums, games, movies, TV shows, DVDs, and formerly, books. For each product, a numerical score from each review is obtained and the total is averaged. It was created and founded by Jason Dietz, Marc Doyle, and Julie Doyle Roberts in 2001. An excerpt of each review is provided along with a hyperlink to the source. Three colour codes of Green, Yellow and Red summarize the critic's recommendation. This gives an idea of the general appeal of the product among reviewers and, to a lesser extent, the public.

The site is somewhat similar to Rotten Tomatoes, but the scoring results sometimes differ very drastically, due to Metacritic's method of scoring that converts each review into a percentage that the site decides for itself, before taking a weighted average based on the critic's fame or stature, and listing different numbers of reviews. Also, Rotten Tomatoes only reviews movies, while Metacritic reviews music albums, video games, movies, DVDs, and TV shows.

Many review websites give a review grade out of five, out of ten, out of a hundred, or even an alphabetical score. Metacritic converts such a grade into a percentage. For reviews with no explicit scores (for example, The New York Times reviews), Metacritic manually assesses the tone of the review before assigning a relevant grade. Weighting is also applied to reviews—those from major periodicals may have a greater effect on the average than niche ones, although Metacritic refuses to reveal what weights are applied to which publications."

As one can see, the background on Wikipedia is suggestive on a number of things related to Metacritic.

Metacritic:

- **Converts each review into a percentage the site decides for itself**
- **Manually assesses the tone of reviews that have no explicit score and then assigns a quantitative score**
- **Refuses to reveal what weights are applied to which publications**

Now, assuming the above is correct, this would make the data shown at Metacritic highly biased and far from the truth. Statistics dictate that you cannot apply scores at your own leisure based on subjective perception or using unvalidated methods and then call that a sophisticated meta-analysis. Therefore, it is time someone takes a closer look at the data at *Metacritic*, in a descriptive way, to explore the data and see what comes up. I have chosen the *Computer Games* listed there for my target of research, as this would also enable some comparison to my previously published *Mobygames: Quantify Me* paper. The objective of my current analysis is to visualize the data, compare critic scores with user scores, in time and cross-sectional, look for patterns and then try to

explain some of those patterns. I have one primary hypothesis, and that states that both critic metascores and user metascores differ significantly, based on the idea that users rate games differently than the “professional reviewers”, especially considering the scoring method of Metacritic is suggestive of statistical flaw. As secondary objectives I would like to look at confounders of scores, such as time, specific critics (or reviewers) and other variables, and comparisons of game release dates with Mobygames data. Finally, some post-hoc exploratory analyses may be warranted as the interpretation of the data is moving forward.

2. Methods

Between 25th of August 2013 and the 10th of September 2013 a custom made tool called *Slurp!* extracted the following data from the Metacritic website (<http://www.metacritic.com>). For each game it collected the Title, Publisher, Developer, Release Date, Platform, Critic Metascore, Number of Critics, User Metascore, Number of Users, Individual Critic Reviews (including Critic Name and links to the original reviews if present), and Individual User Reviews (including User Name and review text of user). This data was then fed to *IBM Statistics 20.0* for further analysis. Descriptive statistics were used to present the data. *User scores* listed at *Metacritic* were multiplied by 10 to obtain a 0-100 range. Kolmogorov-Smirnov Test for Normality was performed to assess the distribution of data. *Critic scores* and *User scores* were considered to be dependent samples, since a. they both came from *Metacritic*, b. they revolve around the same topics (games), c. they may be influenced by each other. For this reason, a Wilcoxon Signed Rank test was chosen to assess statistical differences between these two score types.

3. Results

3.1 Baseline data

The total number of games listed at *Metacritic* was 136.496 at the time of database completion. Game review scores from 14 different platforms were collected (Table 1). While the Critic scores all have a value of 0-100, the User scores have a value of 0-10. Table 1 lists the median scores, as both the *Critic score* and *User score* data was *skewed* and not distributed normally according to the Kolmogorov-Smirnov Test for all platforms. Also, the Valid N variable indicates the number of games where an actual score (Critic or User) was listed at the website out of a total number of games (Total N) registered at the site.

Table 1. Overview of baseline data collected at *Metacritic*

Platform	Audience	Median	5%	95%	Valid N	Total N	Valid %
3DS	Critics	68	43	85	217	423	51%
	Users	7.2	3.7	8.7	184	423	43%
DS	Critics	66	38	83	989	2280	43%
	Users	7.7	3.6	9	590	2280	26%
GBA	Critics	68	40	85.5	600	772	78%
	Users	8.2	4.3	9.5	298	772	39%
iOS	Critics	75	56	88	2088	112243	2%
	Users	7.6	4.5	9.4	413	112243	0%
PC	Critics	72	44	88	3062	9568	32%
	Users	7.5	3.8	9	3647	9568	38%
PSOne	Critics	71	35	92	256	289	89%
	Users	8.5	5.8	9.4	173	289	60%
PS2	Critics	69	45	88	1532	1828	84%
	Users	8.1	4.8	9.2	1339	1828	73%
PS3	Critics	72	44	88	1290	2306	56%
	Users	7.1	3.7	8.5	1283	2306	56%
PSV	Critics	74	49	87	123	333	37%
	Users	7.3	3.8	8.8	154	333	46%
PSP	Critics	69	47	84	605	998	61%
	Users	7.8	4.9	9.1	476	998	48%
Wii	Critics	65	35	84	874	1981	44%
	Users	7.4	3.3	8.9	742	1981	37%
Wii U	Critics	76	40	86	63	107	59%
	Users	6.95	2.3	8.3	72	107	67%
Xbox	Critics	71	43	89	837	860	97%
	Users	8	4.5	9.2	615	860	72%
Xbox 360	Critics	70	41	88	1686	2367	71%
	Users	7.1	3.7	8.6	1628	2367	69%

For example, *Metacritic* listed 112243 games for iOS, while only 413 of those have any user score (0.36%). 5% and 95% percentiles of the scores are also shown. Note that the score of the users is at a range of 0-10, while those of critics are actual different types of scores (e.g. A-F, 0-10, 0-100) converted to a 0-100 range. For statistical purposes the *User* scores have been multiplied by 10 to get a 0-100 range, comparable with the *Critic* scores.

3.2 Critic scores (CS) versus User scores (US)

For each platform, it is possible to calculate whether the differences in *median Critic* and *User* scores are significantly different (**Figure 1**). For each platform there is a significant difference between the median User scores and median Critic scores, except the 3DS, Playstation Vita and Xbox 360. Overall, disregarding platforms, the distribution of differences between both scores is depicted in **Figure 2**.

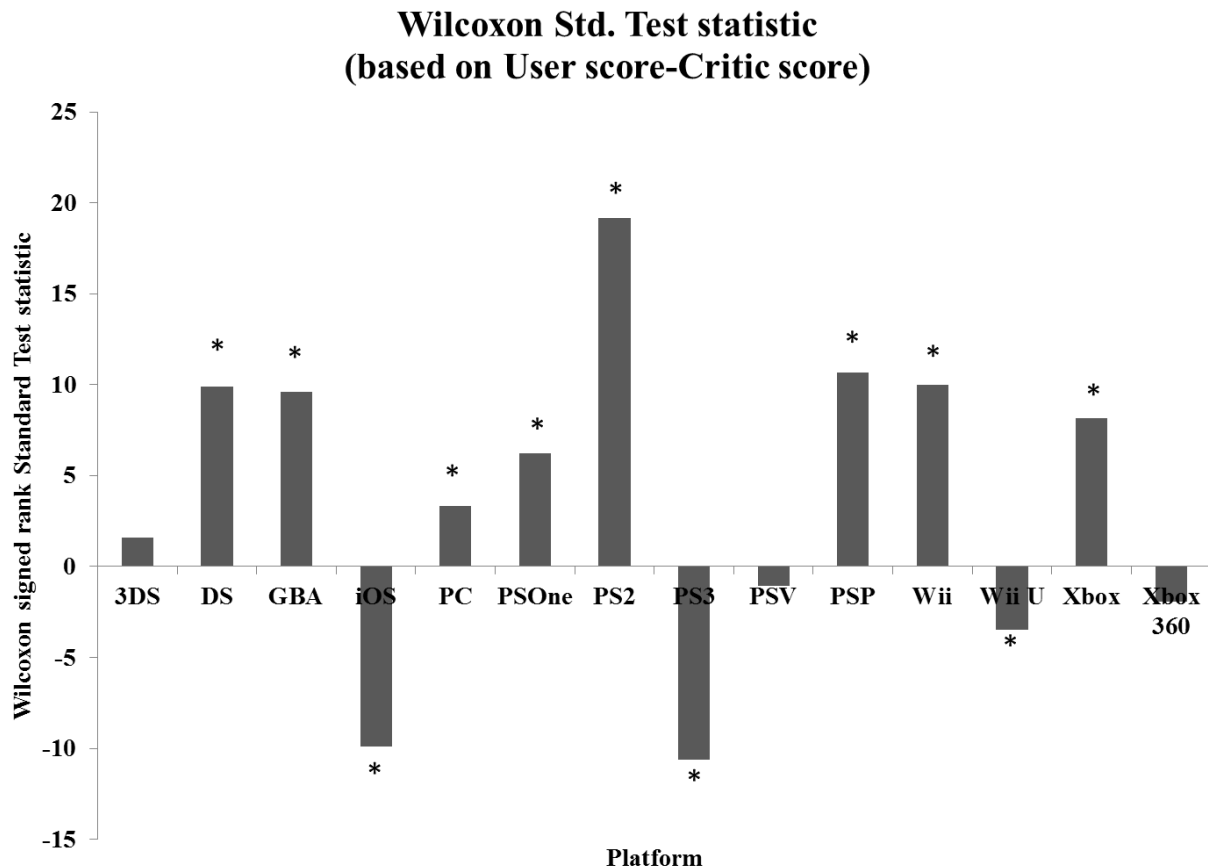


Figure 1. *Wilcoxon Signed Tank test statistic of User score – Critic score by platform at Metacritic.* Asterix indicates a significant difference between User scores and Critic scores (alpha = 0.05).

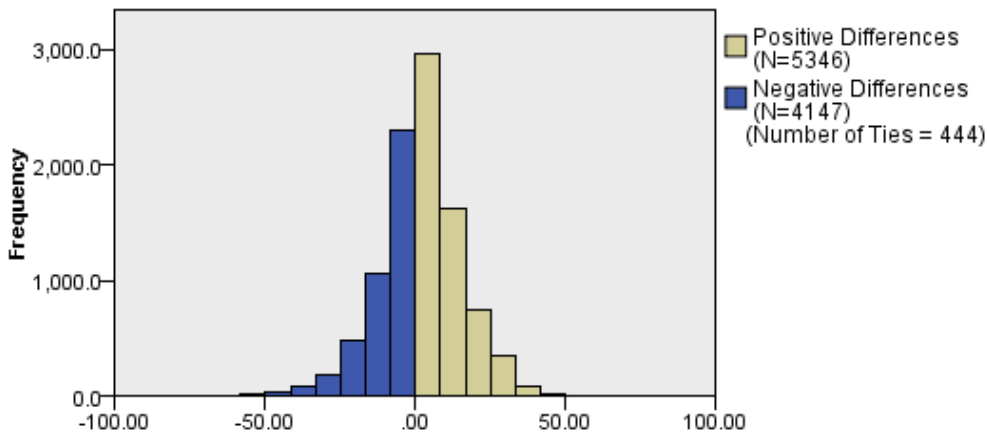


Figure 2. Frequency of differences (User score – Critic score) for all Metacritic entries. Out of 9937 valid comparisons, 5346 had a user score higher than the critic score (54%), 4147 had a user score less than the critic score (41%) and 444 entries shows no difference (5%).

The data illustrated in the figure is informative. Overall, users rated 54% of games higher than the critics, rated 41% of games less than the critics, and showed agreement with critics in 5% of the cases. In all, User and Critic scores differ significantly from each other (standardized test statistic: 14.623, $P < 0.0001$). However, as in **Figure A1** (Appendix) is shown there is a correlation between the number of Critic scores and User scores per game.

3.3.1 Critic scores, user scores in time

Since *Metacritic* keeps track of release dates of each game, we can take a look at the Critic and User scores for these games, per release year. **Figure 3** depicts the median of CS and US by release year, along with their respective 25% and 75% percentile values. The median of the US starts above 90 in 1994 but shows steady decline to 70 in 2012. The median CS is more or less steady for games released between 1994 and 1999 (around 87), but shows a dramatic drop to ~70 for games released after 1999 in 2000 and beyond. This is most probably caused by a *regression to the mean effect*, as there is a large increase in number of reviews from 2000 onwards (i.e. 38 in 1999, to 297 in 2000: year 1994 through 1999 there were a total of 118 critic scores listed). CS median declines further to 67 between 2005 and 2008, when the CS seems to rise again to about 74 in 2012. Especially for games released between 2000 and 2010, CS and US seem to differ, as there is no overlap in either median or 25% and 75% percentiles. From 2010 onwards there is again overlap, but this is driven by the increase in CS. In a univariate linear model, the year of release is significantly associated with User score ($P < 0.0001$), with a B of -1.036 and R^2 of 6.3%.

The median CS by platform also differ (**Figure 4**). In Figure 4 the median CS by platform are shown, along with their 25% and 75% percentiles. The table in **Figure A2A** gives insight into which differences in CS between the platforms are significant. For example, the lowest median CS of 65 given to Wii games is statistically different from all other platforms, except perhaps the 3DS given an alpha of 0.05. That implies that the “population” of critics at *Metacritic* rate Wii games the lowest of all. In contrast, iOS

games seem to be rated among the highest of all, as the CS differs from all other platforms favorably, except perhaps the Wii U and PS Vita, but that is probably due to

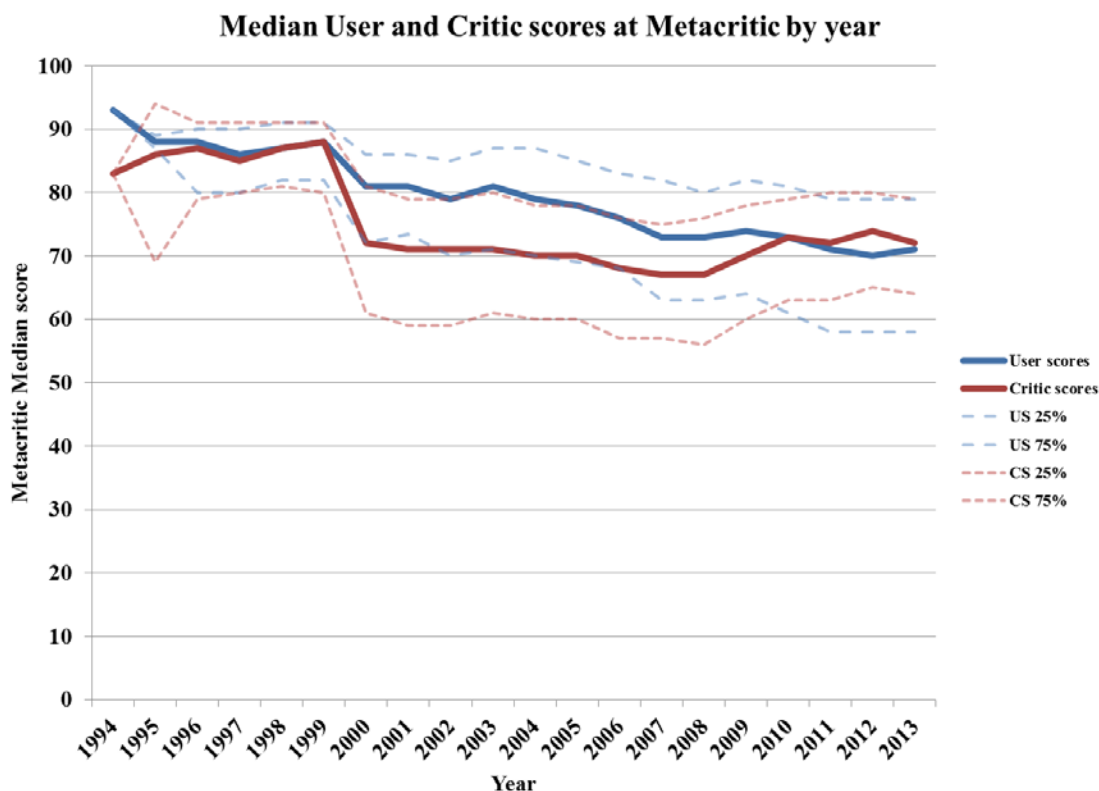


Figure 3. Median User and Critic scores by release year of games. Also shown are their respective 25% and 75% percentiles.

the low number of games listed for Wii U and PS Vita. Between the iOS games and the Wii there is a full point difference in median CS: 76 for the iOS games versus 65 for the Wii games. Likewise, it is possible to do the same exercise for US (**Figure 5**). User scores for Playstation 1 games are rated the highest versus all other platforms (**Figure A2B**), while Wii games are rated the lowest by users. Also high are Gameboy Advance and Playstation 2 games, while on the lower side we find Xbox 360, Playstation 3 and 3DS. On the whole, the data suggests games on older platforms are rated higher by users than games for the new platforms.

The number of critic ratings and user ratings are also different between platforms (**Table 2**). Xbox 360 games have in total 56039 critic ratings, followed by PC games (52796) and Playstation 2 (35320). Wii U games (1482), PSOne (2855) and PSV (3057) are among the least rated by critics. However, these are absolute numbers, and there are large differences in the number of games listed at *Metacritic* by platform. Therefore, it might be more appropriate to look at the median number of ratings per game. Thus, Xbox360 games are rated by 24 critics and 24 users “half” of the time, and iOS games have a median of 2 CR and 7 UR. Overall, the median of ratings at *Metacritic* is 7 (out of 292067 ratings) and 15 (out of 814398 ratings) for CR and UR respectively.

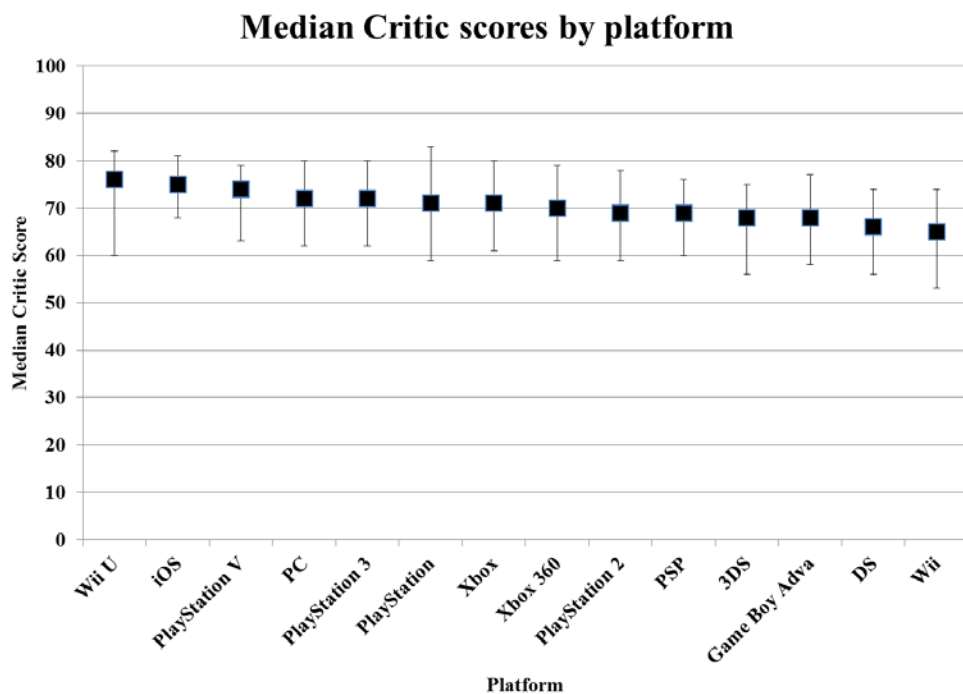


Figure 4. Median Critic scores by platform. Shown are also the 25% and 75% percentiles.

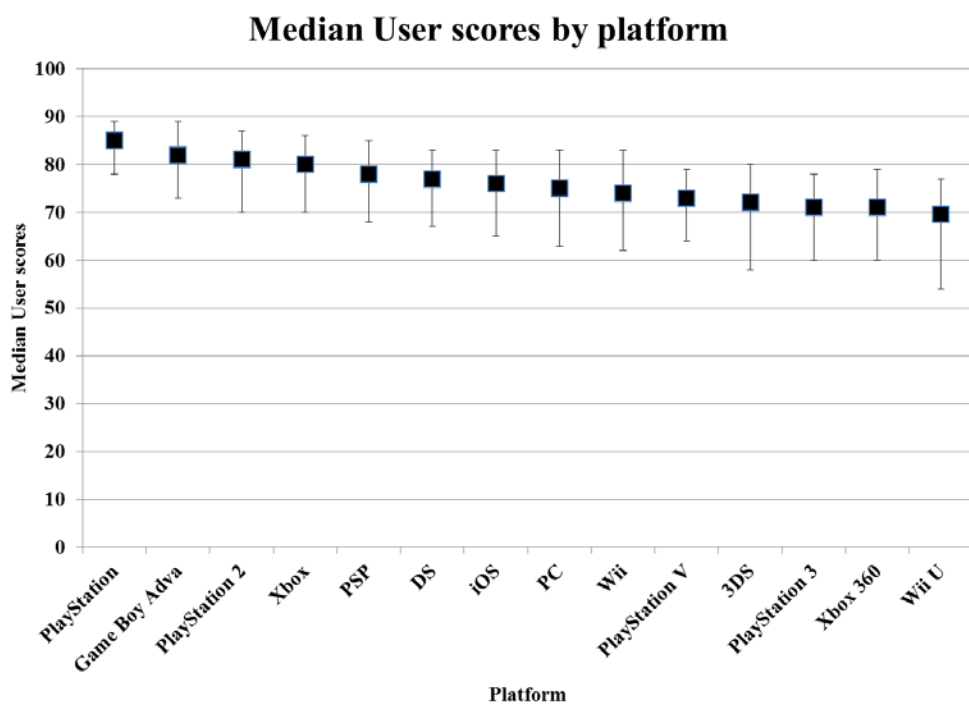


Figure 5. Median User scores by platform. Shown are also the 25% and 75% percentiles.

Table 2. Median and cumulative Critic ratings (CR) and User ratings (UR) by platform.

Platform	Games	Median CR	CR Sum	Median US	UR Sum
<i>Xbox 360</i>	2367	24	56039	24	157534
<i>Xbox</i>	860	23	21936	10	11988
<i>PS2</i>	1828	19	35320	14	37723
<i>PS3</i>	2306	18	35152	22	126749
<i>Wii U</i>	107	14	1482	20	3438
<i>PSP</i>	998	12	13314	12	10419
<i>PC</i>	9568	9	52796	19	397482
<i>PSOne</i>	289	9	2855	14	5802
<i>GBA</i>	772	8	7994	8	4561
<i>3DS</i>	423	7	4973	17	7931
<i>PSV</i>	333	7	3057	20	6979
<i>Wii</i>	1981	7	16897	11	25520
<i>DS</i>	2280	4	18320	10	13199
<i>iOS</i>	112243	2	21932	7	5073

3.3.2 Number of Critic and User reviews in time

For each game listed at *Metacritic* there is a number of Critic reviews and a number of User reviews. First, it is possible to take a look at the number of games listed by release year (as indicated by *Metacritic*). See **Figure 6**.

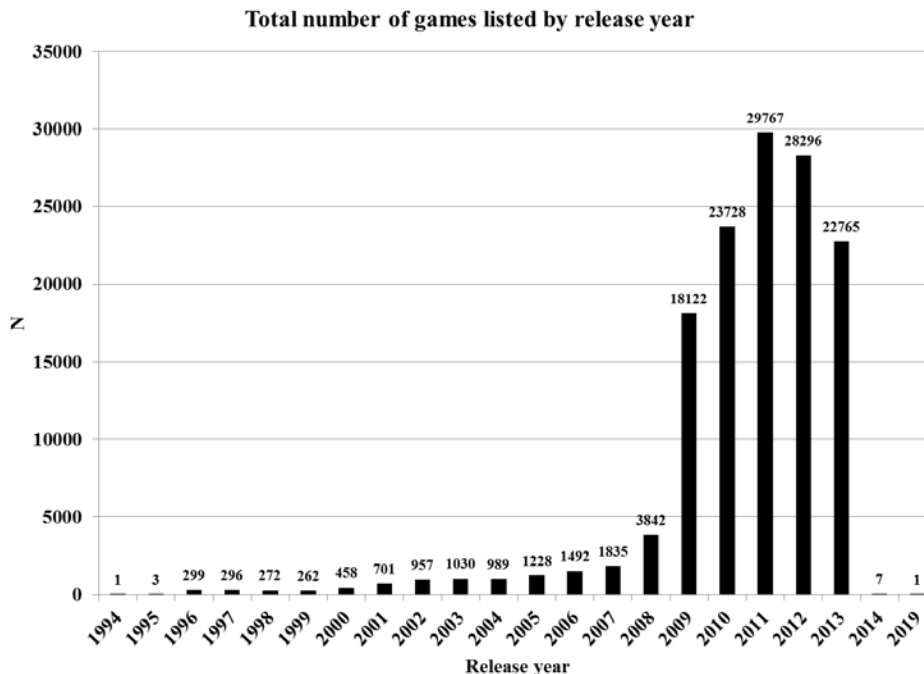


Figure 6. Number of games at Metacritic by release year. 1996 is the year that kicks off.

From the figure, it is evident that games with a release date from 1996 through 2013 are listed, while few from 1994 and 1995. Oddly, 8 games are listed with future release dates (7 in 2014, 1 in 2019). See also **Figure A5** and **A6** to compare with *Mobygames*.

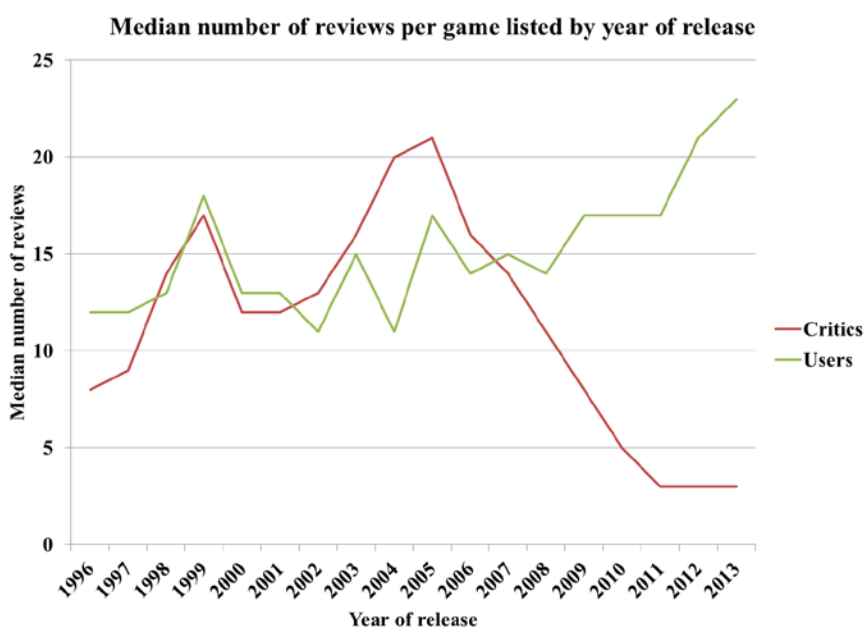


Figure 7. Median number of reviews per game by year of release. Two peaks are visible for the median number of Critic reviews: 1999 and 2005. The data included the value 0 for games for which no Critic review was observed (similar for User reviews).

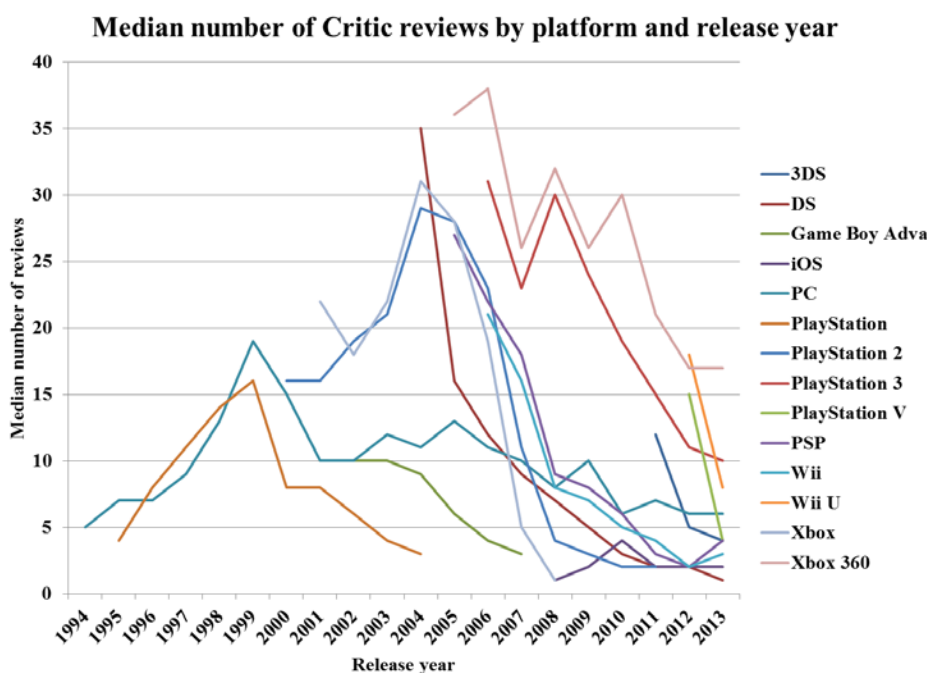


Figure 8. Median number of Critic reviews by platform and release year.

The number of reviews per game (both Critic and User reviews) can be plotted against release year. In **Figure 7** this is done using the median number of reviews. It shows a dramatic drop in median number of Critic reviews from 2006 onwards. As is apparent from **Figure 8**, this seems to be mainly caused by a drop in median number of Critic reviews for Xbox and Playstation 2 games. Of note, those platforms, similar to Playstation One and PC games, seem to have a bell shaped curve, while all other newcomers (e.g. DS, PSP, Wii, Xbox360, Playstation 3) start their curves at the highest median number of Critic reviews and then this value goes down. Taken together, that will cause the dramatic decline as seen in Figure 7, not helped by the fact that iOS games are listed in great numbers, but relatively hardly reviewed (Table 1) at *Metacritic*. Take a look at **Figure A4** to see the same for user reviews.

3.4 Visual plot of differences between US and CS

When we take 10% percentiles of the CS, and then apply those ranges in scores also to the US we can plot the distribution of the number of times each scored review is in either one of those categories. In **Figure A3** the example is presented where the overlap between CS and US is complete and there is no difference in distribution between those 10% percentile ranges in terms of US and CS. However, in **Figure 9** the actual differences are shown. For example, when looking at the lowest 10% percentile of the CS (≤ 50) and plotting the distribution of the US associated with that CS range, it is clear that 50% of the US match this range: ≤ 50 . Yet, there were also quite some US that did not rate the game ≤ 50 , but differently. So 50% of average user scores agreed with that ≤ 50 score, but half also did not. For the highest Critic scores (86+) only 35% of User scores matched, and 65% of average User scores were in a lower range. Overall, from the CS range of 51-58 onwards the number of US that score the games higher than the CS goes from about 65% to about 24% at a CS range of 81-85. It is evidence of Critic scores being much lower than User scores across the board, but more specifically of games scoring very low Critic scores that Users seem to not agree with at all.

Taking a closer look at the top 20 games ranking highest by *Critic score* (**Table 3**) just as an example, for 10 out of those 20 games there is a strong deviation between the average *Critic score* and the average *User score* (>10 points). In each case, the Critics give those games a much higher score than the users do. It is important to note that the median number of User reviews per game is 15 (25% percentiles are ≤ 7 , 8-15, 16-39, 40+ respectively). The number of games for which at least 100 users have scored is 1304 out of 11615 games listed with user review scores. Only 269 games had at least 500 users that scored, and 123 games had at least 1000 user scores. The games listed in Table 3 all have at least 20 user scores, 14 have more than 500 and half (10) have more than 1000. This provides a sense of representation of the true population score.

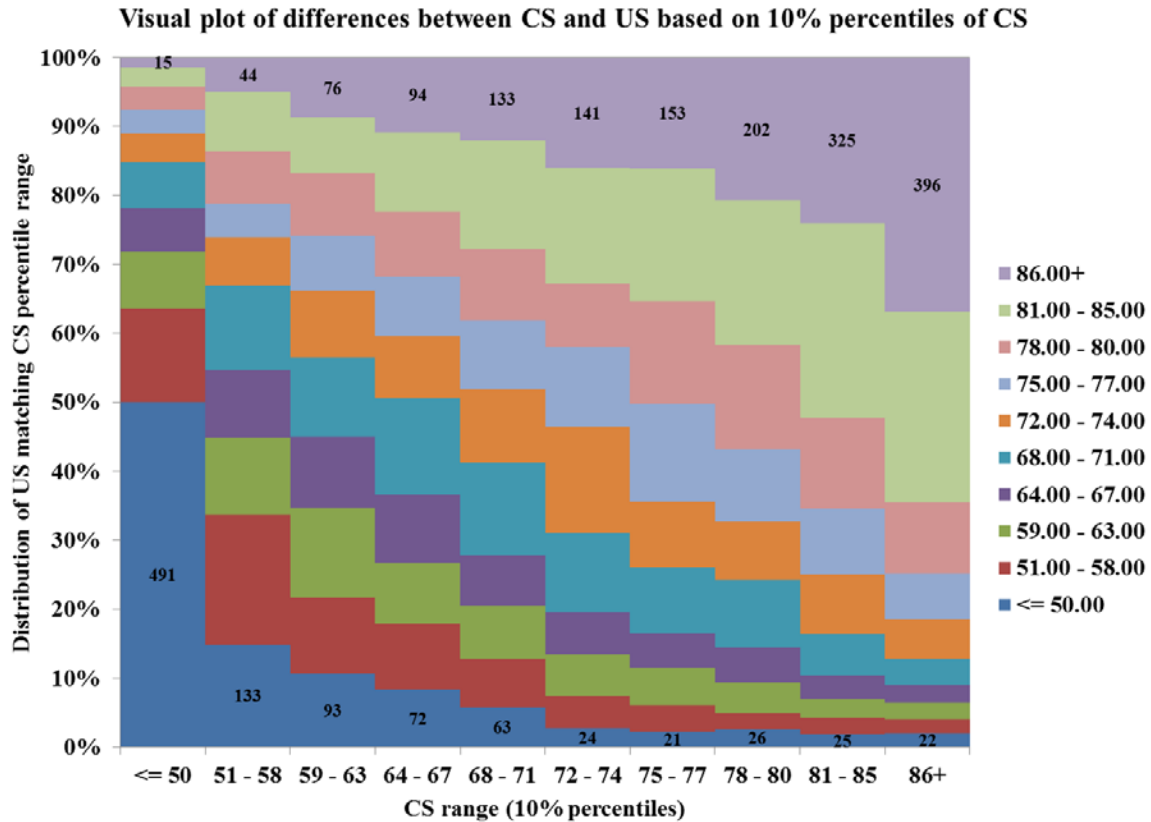


Figure 9.

Table 3. Top 20 highest ranking games at Metacritic by Critic score (CS).

Rank	Name	Publisher	Released	Platform	CS	(n)	US	(n)
1	Grand Theft Auto IV	Rockstar	4/29/2008	PS3	98	64	80	1870
2	Tony Hawk's Pro 2	Activision	9/20/2000	PS	98	19	80	60
3	Grand Theft Auto IV	Rockstar	4/29/2008	Xbox 360	98	86	80	2395
4	Grand Theft Auto III	Rockstar	10/22/2001	PS2	97	56	90	415
5	Tony Hawk's Pro 3	Activision	10/28/2001	PS2	97	34	70	169
6	Halo: Combat Evolved	Microsoft	11/14/2001	Xbox	97	68	90	726
7	Super Mario Galaxy	Nintendo	11/12/2007	Wii	97	73	90	1419
8	Super Mario Galaxy 2	Nintendo	5/23/2010	Wii	97	88	90	1282
9	Super QuickHook	Rocketcat	6/17/2010	iOS	96	9	60	20
10	World of Goo	2D Boy	4/14/2011	iOS	96	6	70	56
11	World of Goo HD	2D Boy	12/16/2010	iOS	96	11	70	45
12	Batman: Arkham City	Warner Bros.	10/18/2011	PS3	96	42	80	1219
13	Uncharted 2: AT	SCEA	10/13/2009	PS3	96	105	90	3486
14	Metal Gear Solid 2	Konami	11/12/2001	PS2	96	42	90	558
15	Resident Evil 4	Capcom	10/25/2005	PS2	96	38	90	528

16	<i>BioShock</i>	2K Games	8/21/2007	PC	96	44	80	2149
17	Half-Life	Sierra	10/31/1998	PC	96	24	90	2021
18	Half-Life 2	VU Games	11/16/2004	PC	96	81	90	6117
19	<i>Out of the Park 2007</i>	<i>Sports I.</i>	3/23/2006	PC	96	5	30	553
20	The Orange Box	EA Games	10/10/2007	PC	96	34	90	1110
					96.5	46.4	80	1309.9

CS: Critic score; US: User score; n: number of reviews. Italics: > 10 points difference between CS and US.

3.5 Quality analysis: 123 games with 1000+ user scores

Though the value of 1000 is an arbitrary number to use to select the 123 games that have at least that number of user scores as input in the average user rating, it may be helpful to shed some light on what is underlying the proven overrating of games by critics, and also the underrating of games by critics. It gives a selection bias if we do only an analysis on those 123 games, yet a game that gets scores from 1000+ *Metacritic* users must be of some importance and it will be interesting to take a closer peak.

In **Figure 10**, the lines depict the average *User scores* of those 123 games, ranked descending for the User scores, and the associated average *Critic scores*. Of interest, while the curves seem to match more or less at the higher scoring games, from about 80% scores the curves start to deviate, when users score games much lower than the critics do. While critic scores hardly go below 70% (average 88%), User scores (average 72%) drop down to 15%. It is therefore of some concern to look closely at the 123 games and note the top 20 games that are overrated by Critics most when compared to the Users. In **Table 4** this is shown. For example, *Company of Heroes 2* for PC has been scored 80% on average by Critics, while Users rated it 16%. *Metacritic* calls that an “Overwhelming dislike” on the user’s part. It certainly is a big deviation from the Critics’ opinion. The same goes for 8 Electronic Arts games (mostly *Dragon Age II* and *Mass Effect* series games) and 8 Activision releases (mostly about *Call of Duty: MW* and *Black Ops*). It is illustrative of how Critics are in no way relating to what User’s feel when they play an actual game. Perhaps more objective is taking the 75%-100% percentile of games with regards to number of user scores (i.e. 40+ user scores per game or higher). There are 2789 games that have a User score and a review score that fit into this category. In **Table A2** the top 50 publishers are listed that were found in this high quartile, along with the median User scores and Critic scores. Of note, take a look at **Table A1** in the Appendix. It lists the top 30 games by the highest number of users that provided input in the *User metascore* for that game. These data suggest that games like *Company of Heroes 2* (as shown before), *Diablo III*, *COD:MW2&3*, *LittleBigPlanet*, *SimCity* and *Mass Effect 3* have been *seriously* highly overrated, as far as gamers (=Users at *Metacritic*) are concerned.

US and CS for 123 games with 1000+ user scores, by US

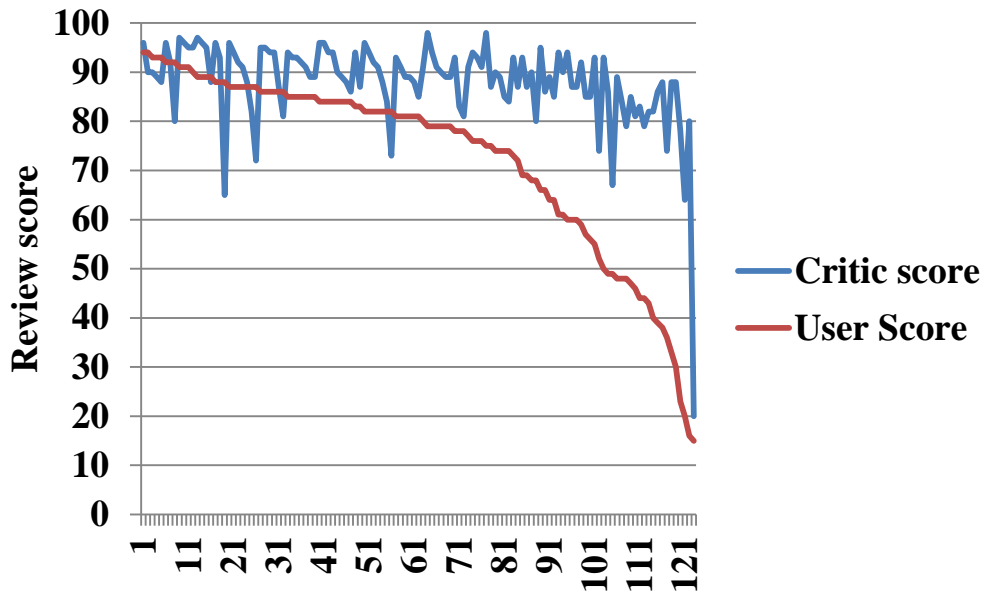


Figure 10. The average user scores for 123 games that are listed at Metacritic with more than 1000 User scores (US). The lines show both the US and the associated Critic scores (CS). The games are ranked by US in a descending manner.

Figure 11 is more illustrative, as it shows the deviations (critic score-user score) from high to low per publisher. Note that positive deviations represent overrating by critics versus users, while negative values would represent underrating by critics. As can be seen, there is more overrating than underrating. For example, Electronic Arts, Activision, Blizzard Entertainment and Rockstar are among those publishers where critics deviate most from users, scoring their games higher than the users.

Table 4. Games listed by Top 20 deviation between average Critic scores and User scores for those games that have at least 1000 user scores.

Name	Publisher	Platform	CS	CS (n)	US	US (n)	Delta
Company of Heroes 2	Sega	PC	80	80	16	4922	64
Call of Duty: Modern Warfare 3	Activision	PS3	88	39	30	4972	58
Call of Duty: Modern Warfare 3	Activision	Xbox 360	88	81	33	8319	55
Call of Duty: Modern Warfare 3	Activision	PC	78	26	23	5275	55
Diablo III	Blizzard	PC	88	86	38	8362	50
Call of Duty: Modern Warfare 2	Activision	PC	86	40	39	5820	47
SimCity	Electronic Arts	PC	64	75	20	4077	44
Mass Effect 3	Electronic Arts	PS3	93	30	50	1500	43
Dragon Age II	Electronic Arts	PS3	82	52	40	1066	42
Mass Effect 3	Electronic Arts	PC	89	23	48	4560	41

Call of Duty: Black Ops II	Activision	Xbox 360	83	73	44	1804	39
Dragon Age II	Electronic Arts	PC	82	45	43	3808	39
Mass Effect 3	Electronic Arts	Xbox 360	93	74	55	3350	38
DmC: Devil May Cry	Capcom	PS3	85	40	47	1092	38
Call of Duty: Black Ops II	Activision	PC	74	11	36	1083	38
DmC: Devil May Cry	Capcom	Xbox 360	86	65	49	1217	37
Spore	Electronic Arts	PC	84	75	48	1422	36
Call of Duty: Black Ops	Activision	PC	81	29	46	1283	35
Dragon Age II	Electronic Arts	Xbox 360	79	75	44	2187	35
Call of Duty: Modern Warfare 2	Activision	PS3	94	67	60	1670	34

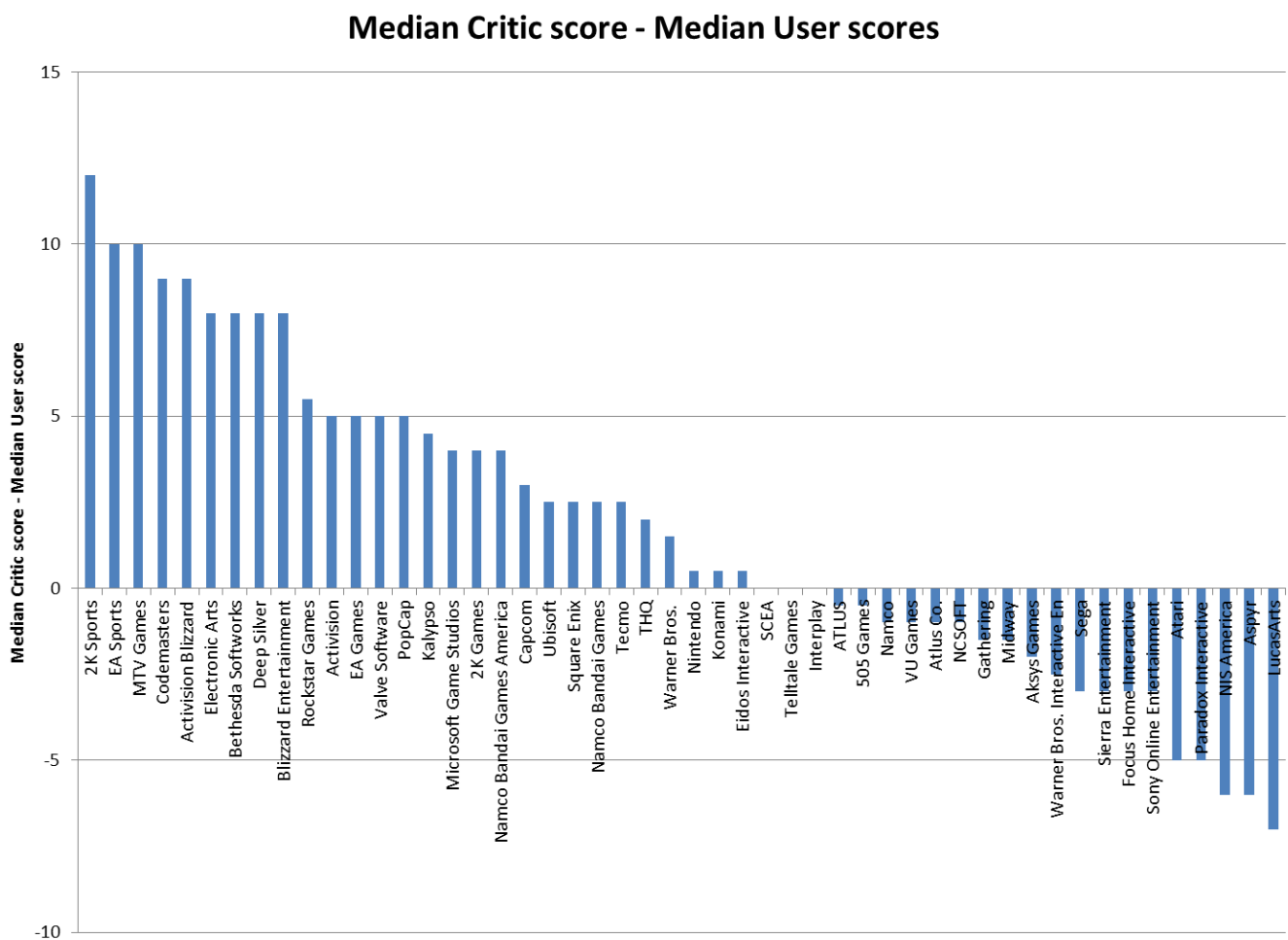


Figure 11. Forest plot of Median Critic score – median User score of top 50 publishers found in highest quartile of number of users that provided input in the User score of a game.

3.6 Critics

It is of further interest to look into the critics themselves. **Table 5** indicates the top 20 of critics listed at *Metacritic* out of 366 unique critics found along with their part of the total of 285341 reviews.

Table 5. Top 20 of game critics listed at *Metacritic*, by number of critic scores.

Rank	Critic	Mean	Median	N	% of total
1	IGN	70 (\pm 17)	73 (60-82)	11011	3.9%
2	GameSpot	68 (\pm 15)	70 (60-80)	8767	3.1%
3	GameZone	74 (\pm 14)	76 (68-85)	6151	2.2%
4	Cheat Code Central	72 (\pm 16)	74 (60-84)	5520	1.9%
5	Game Informer	74 (\pm 15)	78 (65-85)	5113	1.8%
6	GameSpy	71 (\pm 17)	75 (60-84)	4416	1.5%
7	Eurogamer	67 (\pm 18)	70 (60-80)	4159	1.5%
8	GamePro	76 (\pm 16)	80 (70-90)	4128	1.4%
9	Worth Playing	72 (\pm 15)	75 (65-85)	4020	1.4%
10	Gaming Age	70 (\pm 21)	75 (58-83)	3545	1.2%
11	Game Revolution	65 (\pm 21)	75 (50-83)	3411	1.2%
12	1UP	70 (\pm 19)	75 (58-83)	3334	1.2%
13	Game Chronicles	77 (\pm 15)	80 (69-88)	2904	1.0%
14	Pocket Gamer UK	68 (\pm 14)	70 (60-80)	2893	1.0%
15	GamesRadar	68 (\pm 17)	70 (60-80)	2771	1.0%
16	Games Master UK	73 (\pm 15)	77 (65-84)	2747	1.0%
17	148Apps	77 (\pm 10)	80 (70-80)	2553	0.9%
18	Game Over Online	77 (\pm 13)	80 (70-86)	2519	0.9%
19	G4 TV	64 (\pm 20)	60 (60-80)	2495	0.9%
20	GameShark	70 (\pm 21)	75 (58-83)	2428	0.9%

Mean (std. dev.) critic scores, median (25%-75%) critic scores are shown.

In addition, in **Figure 12**, the ratio of number of reviews divided by number of games per platform is shown. Xbox games have approximately 25 reviews (by Critics) per game at *Metacritic*, while PC games and iOS games have the lowest ratio (5.5 and 0.1 review per game respectively).

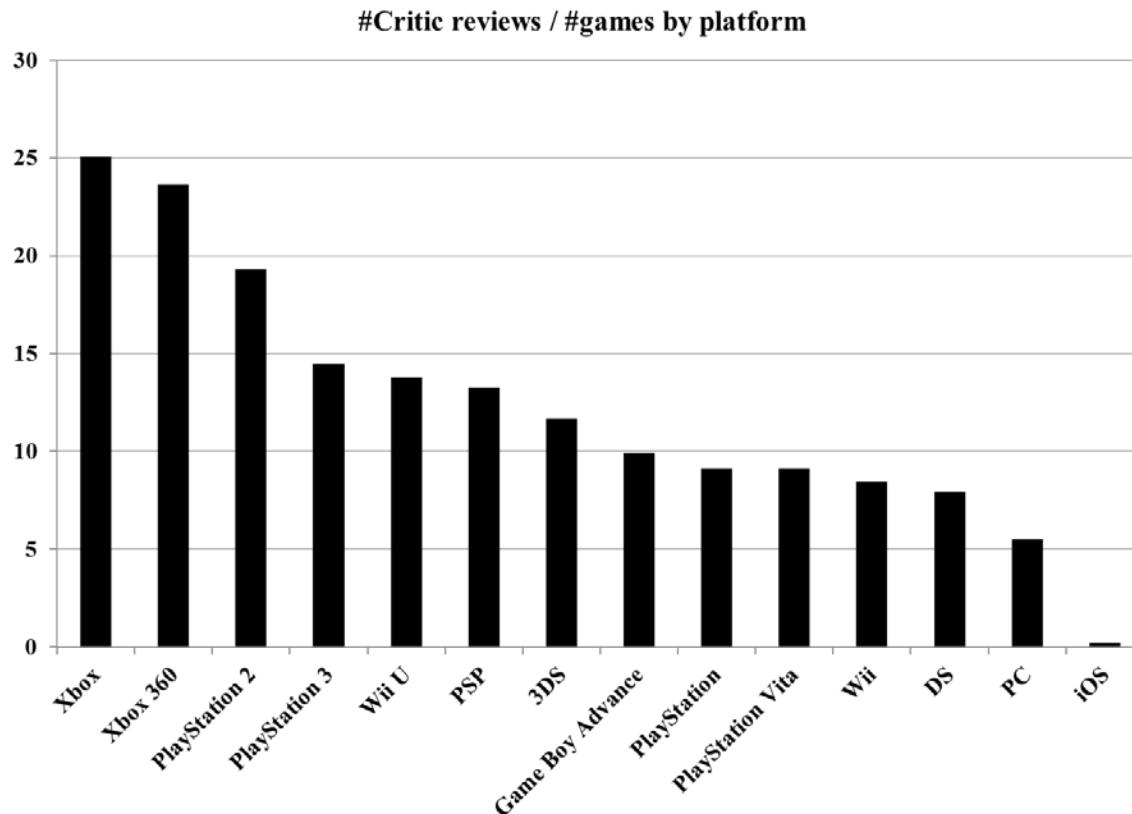


Figure 12. Number of critic reviews divided by number of games by platform.

4. Discussion

The current novel analysis of the data collected by *Metacritic* shows for the first time the quantity and some of the quality of the *metascores*. The results indicate that critic metascores deviate statistically from user metascores. Overall, user metascores are higher than critic metascores across platforms, but when looking at the games receiving high scores from the critics the majority of the users score these games lower, indicating possible bias and overrating on the part of critics. Similarly, games at the lower end of the critic score (scores equal to or less than 50) get higher ratings by users half of the time, suggestive of underrating by critics. The disagreement between users and critics is as follows: users rated 54% of games higher than the critics, rated 41% of games less than the critics, and showed agreement with critics in 5% of the cases, a disagreement of 95% when looking strictly at the absolute metascores. Interestingly, the *Metacritic* database boasts a huge amount of iOS game entries, but collected critic reviews for only 2% of those, and hardly any user reviews. Clearly, *Metacritic* aims to include all apps, but fails to collect reviews, and users can't be bothered. Nevertheless, the critic scores for iOS games are highest among all platforms, statistically different from all other median critic scores (by platform). This in itself is of interest; it might indicate that those critics listed rate iOS apps high in a way much different from critics of games on "true" gaming platforms. The big part of Nintendo platforms, such as the Wii,

the GBA, DS and 3DS do not score well. Games listed at *Metacritic* for those platforms get significantly lower scores by critics, though the Wii U games seem to be on par with the iOS (yet only 107 games for the Wii U were listed). Also, *Metacritic* is heavily focused on the Xbox (and Xbox 360) platform, as well as the Playstation platform, collecting a lot of reviews for these platforms, but much less for other platforms. This is indicative of strong bias, especially where PC games are concerned. PC gaming websites/portals referring to *Metacritic* to aide their users to get an idea how good a game is should probably stop doing this, Steam among those. It's not accurate, incomplete and unfair.

The median user metascores, as well as the median critic metascores (to a lesser extend) decrease with increasing release year of a listed game. As the number of reviews per game seems to increase this effect could in part be explained by a regression to the mean, the more scores collected the closer the metascore gets to the true population mean. However, this doesn't quite hold true for the critic scores. The number of critic scores collected by game release year plummets from 2006 onwards. That is odd. What is happening there? Are there fewer sources for *Metacritic* to collect their critic reviews from? Doubtful. Indeed, it seems to be partly caused by the decline in reviews for Xbox and Playstation 2 games specifically. Surely, those critics moved on to review other games for other platforms, and did not just vanish? It cannot be explained by a drop in number of games. This would introduce bias on the part of *Metacritic*.

A more qualitative analysis of the 123 games that got at least 1000 user scores to offset the user metascore revealed that those games were rated quite highly by the critics. Indeed, when only 123 games out of 136.496 have so many users taking the time to give their score, the games must be special. It is therefore a surprise to find that many of those special high ranking games actually get a low user score. While critic scores hardly went below 70% (average 88%), User scores (average 72%) drop down to as low as 15%. I presented a list of the top 20 games (out of these 123) where the deviation between critic metascore and user metascore was highest. Company of Heroes 2, Call of Duty: Modern Warfare 3 and Diablo III are among those where raving critic scores were debunked by users. One wonders how this can be. How can critics not know that these games are actually not that good, if we assume A. users are right and B. users don't bash a game for good fun? We can understand that if B is apparent, this would introduce a huge bias on the user's part, but the very fact that there is such a large discrepancy between critic scores and user scores throughout the *Metacritic* database is proof enough that critics and users don't see the games the same way. One might assume that users represent the true "end-user" of a game, and as such the customer is always right. Then what are the critics doing?

Taking a peek at the publishers and the deviation between critic metascores and user metascores we see that 2K, Blizzard, Electronic Arts, Bethesda and Rockstar are among those whose games are highly overrated as far as users are concerned. At the other end there is Sierra, Sony and Lucas Arts whose games seem to be underrated by the critics, according to the users. Could hyping by at work here?

The reviews from critics that are listed at *Metacritic* are dominated by those by IGN, GameSpot, GameZone, Cheat code central and Game Informer, and other US

websites. Though that is an obvious (US) bias, it need not necessarily be a big issue. However, with the notion that critics seem to overrate games released by big American publishers, one wonders if there is a connection between reviewers and hyping certain games. And how does that mix with the clear focus on Xbox and Xbox360 reviews? An example of this is the game Metal Slug 3 for the Xbox that listed 56 critic reviews, but only 7 user scores, seems nobody wanted to score that game. On the other hand, Half-Life: Counterstrike got only 11 critic reviews, but 6101 user scores. When games were sorted by number of user scores Counterstrike: Source topped the list with 9471 user scores (and only 9 critic scores).

Limitations

The limitations of the current paper could be that the data was collected over a period of a week, and some new additions for some platforms were not taken into account, due to the technical process of collecting the data. However, these will be very, very small in number.

Future prospects

There are many questions that remain unanswered, as discussed above. It will be worthwhile to take a deeper look at the critics that reviewed all of these games, and what scores they used. In addition, to what extent are certain critics associated with publishers if we compare their ratings across the board? If there such an association at all? And if so, who? And which reviewer is actually always spot on with the final score? This will require a more qualitative process where I look at the actual review data, and note the game website, as well as the individual reviewer.

Conclusion

The data at *Metacritic* leaves much to be desired and seems to be heavily biased. When critic scores do not comply with user scores in the majority of cases, which has been shown in this paper, and the selection of critics is seemingly under no formal control, the validity and accuracy of the data is low. Caution is necessary when using *Metacritic* information to guide your own decision to buy or not buy a game. Until there is more transparency on how this process takes place at *Metacritic*, more transparency on the flow of funding from which sources and the observed biases are removed, the database is of limited use for the end-users.

Appendix

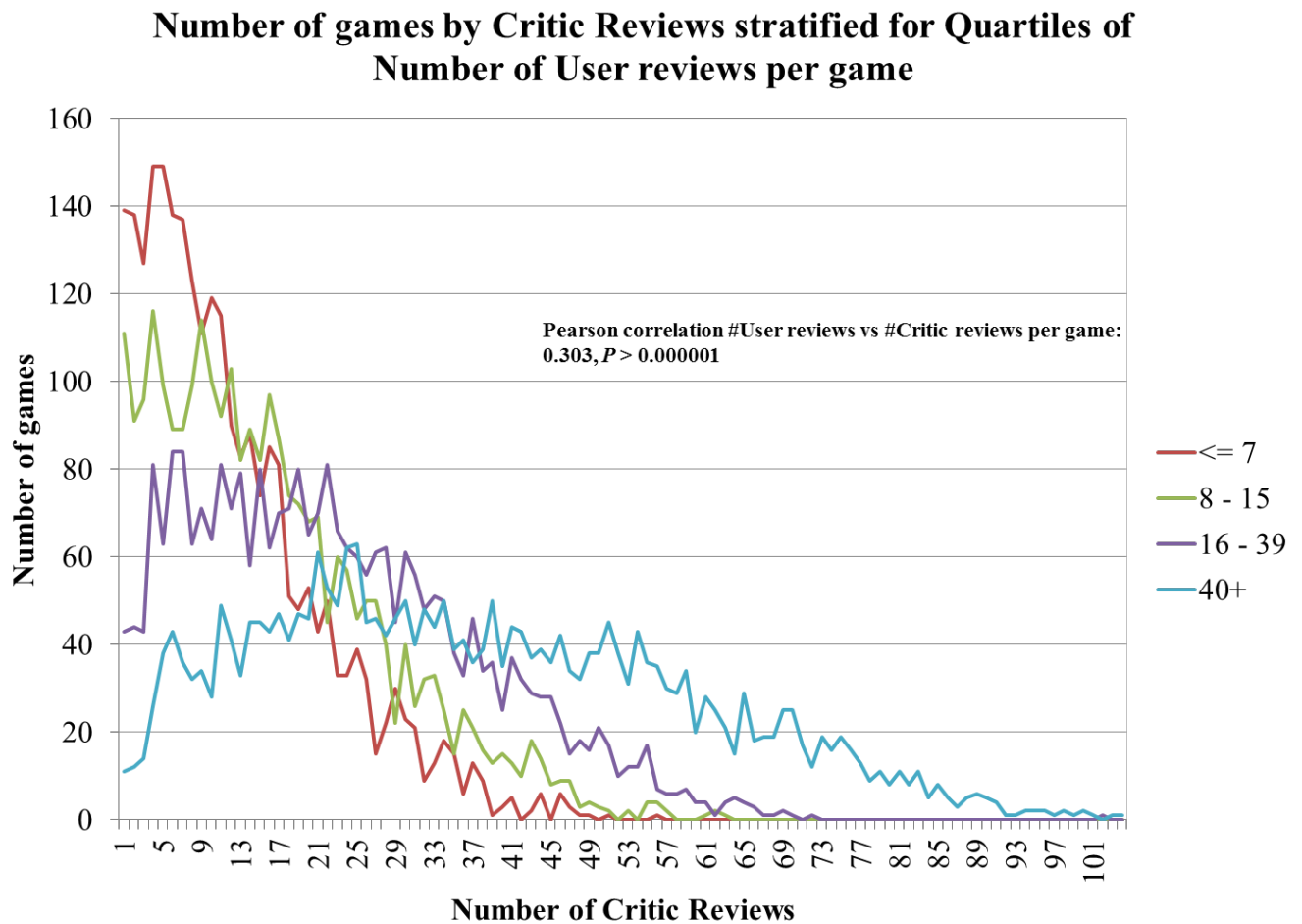


Figure A1. The data are stratified for the quartiles of the number of user reviews per game (legend shows the ranges of the quartiles). It then shows the distribution of the number of games by the number of critic reviews listed per game. Example: the number of games that got 7 or less User reviews AND that got 4 or 5 Critic reviews was about 150 (see red line peak). Interestingly, there are games that got over 50 Critic reviews and only 7 or less user reviews. Indeed, Metal Slug 3, for the Xbox got 56 Critic reviews at *Metacritic* and 7 User scores at the time of database lock (<http://www.metacritic.com/game/xbox/metal-slug-3>). On the other end of the zxspectrum there is Half Life: Counter-strike with 11 Critic reviews and 6101 User scores (<http://www.metacritic.com/game/pc/half-life-counter-strike>).

Table A1. Top 30 games ranked by number of user scores received at *Metacritic*.

Rank	Game title	Critics (n)	CS	Users (n)	US	Delta (CS-US)
1	Counter-Strike: Source	9	88	9471	89	-1
2	Diablo III	86	88	8362	38	50
3	Call of Duty: Modern Warfare 3	49	85	6189	28.67	56.33
4	Half-Life 2	81	96	6117	92	4
5	Half-Life: Counter-Strike	11	88	6101	93	-5
6	Company of Heroes 2	80	80	4922	16	64
7	LittleBigPlanet	85	95	4888	66	29
8	Team Fortress 2	17	92	4629	92	0
9	BioShock Infinite	68	94	4410	85	9
10	Portal 2	52	95	4143	86	9
11	Call of Duty: Modern Warfare 2	70	90	4097	50	40
12	SimCity	75	64	4077	20	44
13	Mass Effect 3	49	91	3955	51.5	39.5
14	Infestation: Survivor Stories	13	20	3517	15	5
15	Uncharted 2: Among Thieves	105	96	3486	87	9
16	Halo 3	86	94	3393	76	18
17	Left 4 Dead	58	89	3334	93	-4
18	The Last of Us	95	95	3330	91	4
19	The Elder Scrolls V: Skyrim	46	94	3224	75	19
20	Portal	27	90	3065	94	-4
21	Dragon Age II	60	81	2998	43.5	37.5
22	Battlefield 3	59	87	2978	74	13
23	The Witcher 2: AOK Kings	76	88	2940	84	4
24	Metal Gear Solid 4: GotP	82	94	2765	87	7
25	Empire: Total War	62	90	2664	68	22
26	Uncharted 3: Drake's Deception	97	92	2445	82	10
27	Minecraft	33	93	2415	73	20
28	Halo 4	87	87	2398	69	18
29	Gears of War 2	90	93	2375	76	17
30	Starcraft II: Wings of Liberty	82	93	2370	81	12

CS: Critic metascore; US: User metascore.

Platform	3DS	DS	GBA	iOS	PC	PSOne	PS2	PS3	PSV	PSP	Wii	Wii U	Xbox	Xbox 3
3DS	-	0.46	0.793	1E-07	8E-05	0.0004	0.202	0.0002	0.02	0.069	0.059	0.001	0.005	0.007
DS	0.46	-	0.005	1E-07	1E-07	1E-07	1E-06	1E-07	1E-06	1E-04	0.008	4E-05	1E-06	1E-06
GBA	0.793	0.005	-	1E-07	5E-06	4E-05	0.22	4E-05	0.004	0.129	2E-04	0.007	0.002	0.009
iOS	1E-07	1E-07	1E-07	-	1E-07	1E-07	1E-07	1E-07	0.052	1E-07	1E-07	0.053	1E-07	1E-07
PC	8E-05	1E-07	5E-06	1E-07	-	0.006	1E-07	0.865	0.502	8E-06	1E-07	0.247	0.672	0.001
PSOne	0.0004	1E-07	4E-05	1E-07	0.006	-	0.002	0.021	0.119	8E-06	1E-07	0.272	0.038	0.003
PS2	0.202	1E-06	0.22	1E-07	1E-07	0.002	-	8E-06	0.008	0.044	1E-07	0.02	0.003	0.023
PS3	0.0002	1E-07	4E-05	1E-07	0.865	0.021	8E-06	-	0.655	2E-05	1E-07	0.327	0.624	0.008
PSV	0.02	1E-06	0.004	0.0517	0.502	0.119	0.008	0.655	-	0.002	1E-06	0.529	0.291	0.164
PSP	0.069	0.0001	0.129	1E-07	8E-06	8E-06	0.044	2E-05	0.002	-	1E-07	0.003	0.001	0.002
Wii	0.059	0.008	2E-04	1E-07	1E-07	1E-07	1E-07	1E-07	1E-06	1E-07	-	7E-05	1E-07	1E-07
Wii U	0.001	3.9E-05	0.007	0.053	0.247	0.272	0.02	0.327	0.529	0.003	7E-05	-	0.194	0.123
Xbox	0.005	1E-06	0.002	1E-07	0.672	0.038	0.003	0.624	0.291	0.001	1E-07	0.194	-	0.362
Xbox 360	0.007	1E-06	0.009	1E-07	0.001	0.003	0.023	0.008	0.164	0.002	1E-07	0.123	0.362	-

Figure A2A. Two-sided P values of Two-Sample Kolmogorov-Smirnov Test comparing Critic scores between two platforms. Yellow denotes a significant deviation.

Platform	3DS	DS	GBA	iOS	PC	PSOne	PS2	PS3	PSV	PSP	Wii	Wii U	Xbox	Xbox 360
3DS	-	0.001	1E-07	0.0055	0.011	1E-07	1E-07	0.309	0.618	6E-06	0.019	0.3	1E-07	0.907
DS	0.001	-	1E-07	0.304	0.01	1E-07	1E-07	1E-07	0.001	0.043	0.001	0.0002	1E-05	1E-07
GBA	1E-07	1E-07	-	2E-06	1E-07	0.03	0.027	1E-07	1E-07	4E-04	1E-07	1E-07	0.006	1E-07
iOS	0.0055	0.304	2E-06	-	0.219	1E-07	1E-07	1E-07	0.003	0.051	0.155	0.0002	0.001	1E-07
PC	0.011	0.01	1E-07	0.219	-	1E-07	1E-07	1E-07	0.012	1E-04	0.408	0.001	1E-07	1E-07
PSOne	1E-07	1E-07	0.03	1E-07	1E-07	-	9E-05	1E-07	1E-07	1E-07	1E-07	1E-07	3E-06	1E-07
PS2	1E-07	1E-07	0.027	1E-07	1E-07	9E-05	-	1E-07	1E-07	0.003	1E-07	1E-07	0.338	1E-07
PS3	0.309	1E-07	1E-07	1E-07	1E-07	1E-07	1E-07	-	0.043	1E-07	1E-07	0.412	1E-07	0.039
PSV	0.618	0.001	1E-07	0.003	0.012	1E-07	1E-07	0.043	-	5E-06	0.059	0.218	1E-07	0.17
PSP	6E-06	0.043	4E-04	0.051	0.0001	1E-07	0.003	1E-07	5E-06	-	5E-05	1E-06	0.288	1E-07
Wii	0.019	0.001	1E-07	0.155	0.408	1E-07	1E-07	1E-07	0.059	5E-05	-	0.004	1E-07	1E-07
Wii U	0.3	0.00016	1E-07	0.0002	0.001	1E-07	1E-07	0.412	0.218	1E-06	0.004	-	1E-07	0.341
Xbox	1E-07	1.1E-05	0.006	0.001	1E-07	3E-06	0.338	1E-07	1E-07	0.288	1E-07	1E-07	-	1E-07
Xbox 360	0.907	1E-07	1E-07	1E-07	1E-07	1E-07	1E-07	0.039	0.17	1E-07	1E-07	0.341	1E-07	-

Figure A2B. Two-sided P values of Two-Sample Kolmogorov-Smirnov Test comparing User scores between two platforms. Yellow denotes a significant deviation.

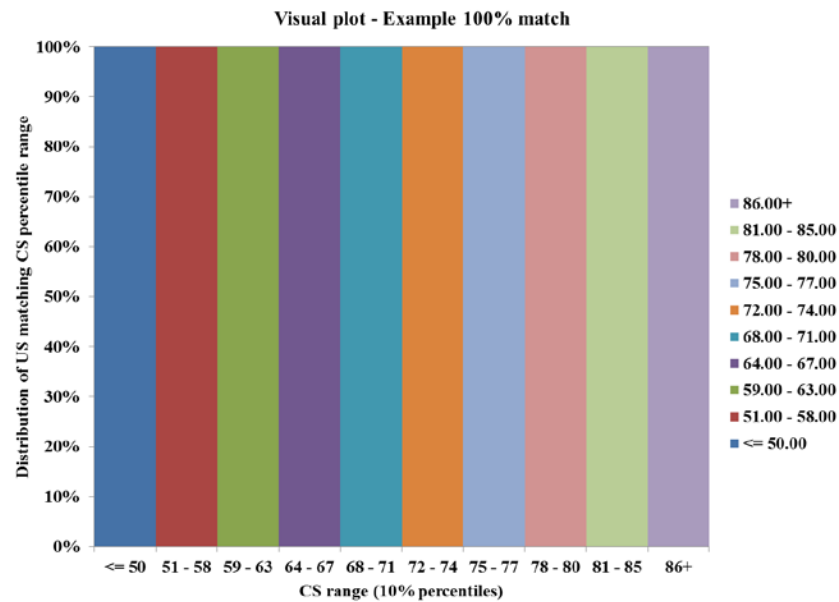


Figure A3. Theoretical visual plot of total agreement between Critic metascores and user metascores.

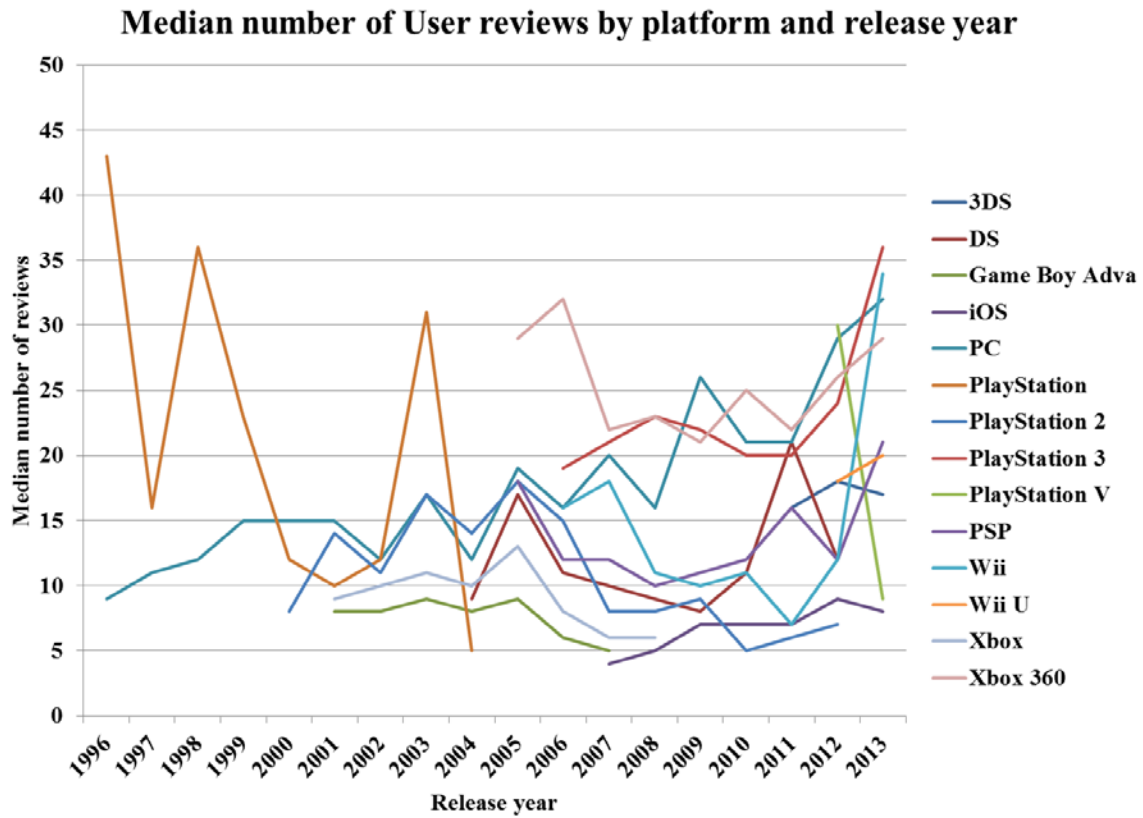


Figure A4. Median number of User reviews by platform by year of release.

Table A2. Top 50 Publishers represented in the 75%-100% percentile based on the number of User scores that gave the Metacritic User rating.

Rank	Publisher	Users	Critics	N	Delta
1	Electronic Arts	73	81	222	8
2	Ubisoft	77.5	80	174	2.5
3	SCEA	83	83	141	0
4	Activision	73	78	136	5
5	Nintendo	84.5	85	118	0.5
6	Capcom	76	79	107	3
7	Microsoft Game Studios	78	82	104	4
8	Sega	77	74	102	-3
9	EA Games	77	82	94	5
10	THQ	78	80	79	2
11	Konami	79.5	80	74	0.5
12	Square Enix	76.5	79	74	2.5
13	Bethesda Softworks	72	80	65	8
14	2K Games	79	83	64	4
15	Atari	79	74	59	-5
16	EA Sports	72	82	51	10
17	Rockstar Games	82.5	88	48	5.5
18	Codemasters	72	81	42	9
19	Eidos Interactive	79.5	80	42	0.5
20	LucasArts	87	80	39	-7
21	Warner Bros.	77.5	79	36	1.5
22	Namco Bandai Games America	74	78	34	4
23	Paradox Interactive	77	72	32	-5
24	Telltale Games	82	82	29	0
25	ATLUS	83.5	83	24	-0.5
26	Namco Bandai Games	73.5	76	24	2.5
27	Deep Silver	63	71	19	8
28	Sierra Entertainment	85	82	19	-3
29	Namco	87	86	16	-1
30	Blizzard Entertainment	81	89	15	8
31	NIS America	72	66	15	-6
32	Valve Software	84	89	15	5
33	2K Sports	75	87	14	12
34	505 Games	72.5	72	14	-0.5
35	VU Games	82	81	14	-1
36	Focus Home Interactive	72	69	13	-3
37	Kalypso	51.5	56	12	4.5
38	Aspyr	75	69	11	-6
39	Interplay	89	89	11	0

40	Activision Blizzard	75	84	10	9
41	Atlus Co.	80	79	10	-1
42	Gathering	86.5	85	10	-1.5
43	Midway	78.5	77	10	-1.5
44	NCSOFT	83	82	10	-1
45	Tecmo	81.5	84	10	2.5
46	Warner Bros. Interactive En	79.5	77	10	-2.5
47	Aksys Games	86	84	9	-2
48	MTV Games	81	91	9	10
49	PopCap	82	87	9	5
50	Sony Online Entertainment	75	72	9	-3

User and Critic scores are listed. N: number of games, delta is deviation between critic scores and user scores. A positive value represent a form of overrating, a negative value an underrating.

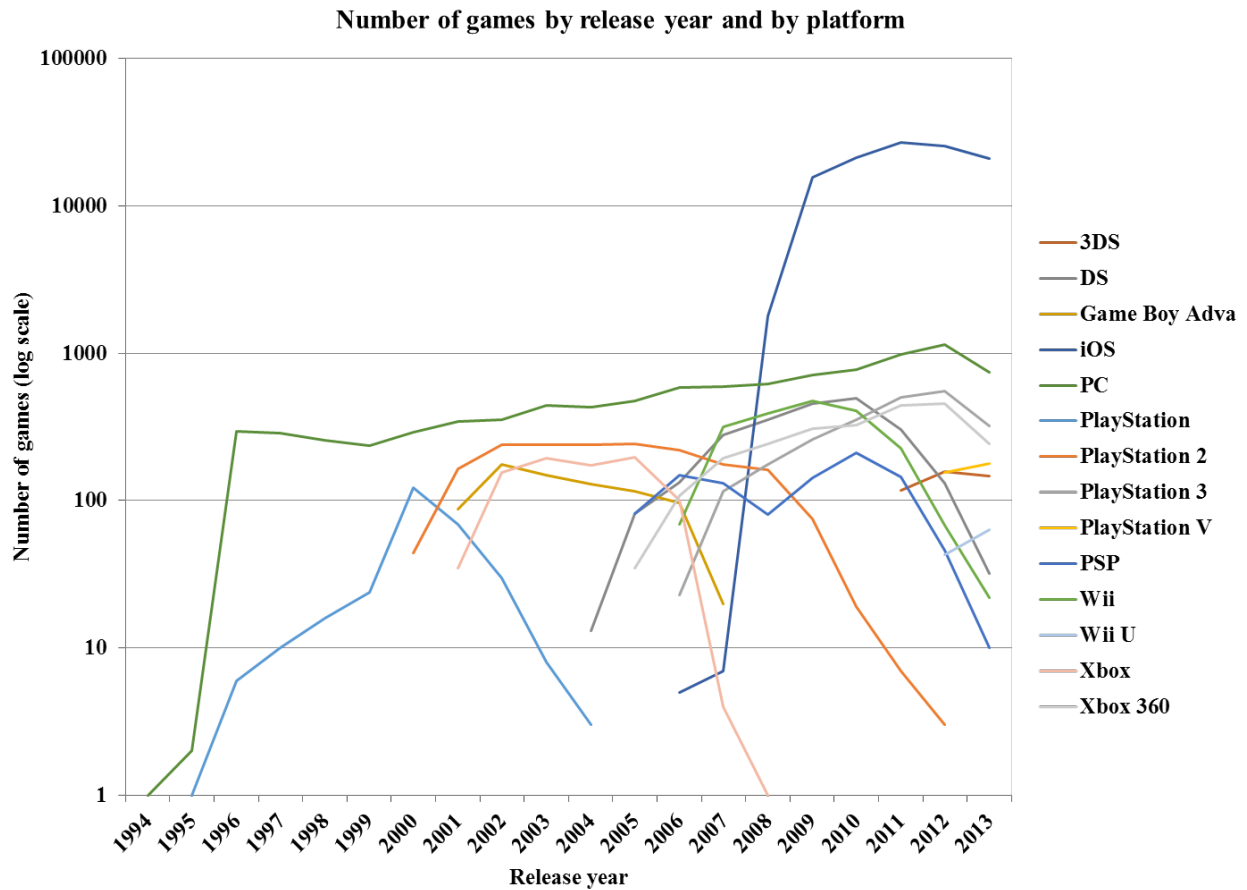


Figure A5. Number of games listed at *Metacritic* by platform and release year. Note the logarithmic scale of the Y axis, necessary because of the large number of iOS games.

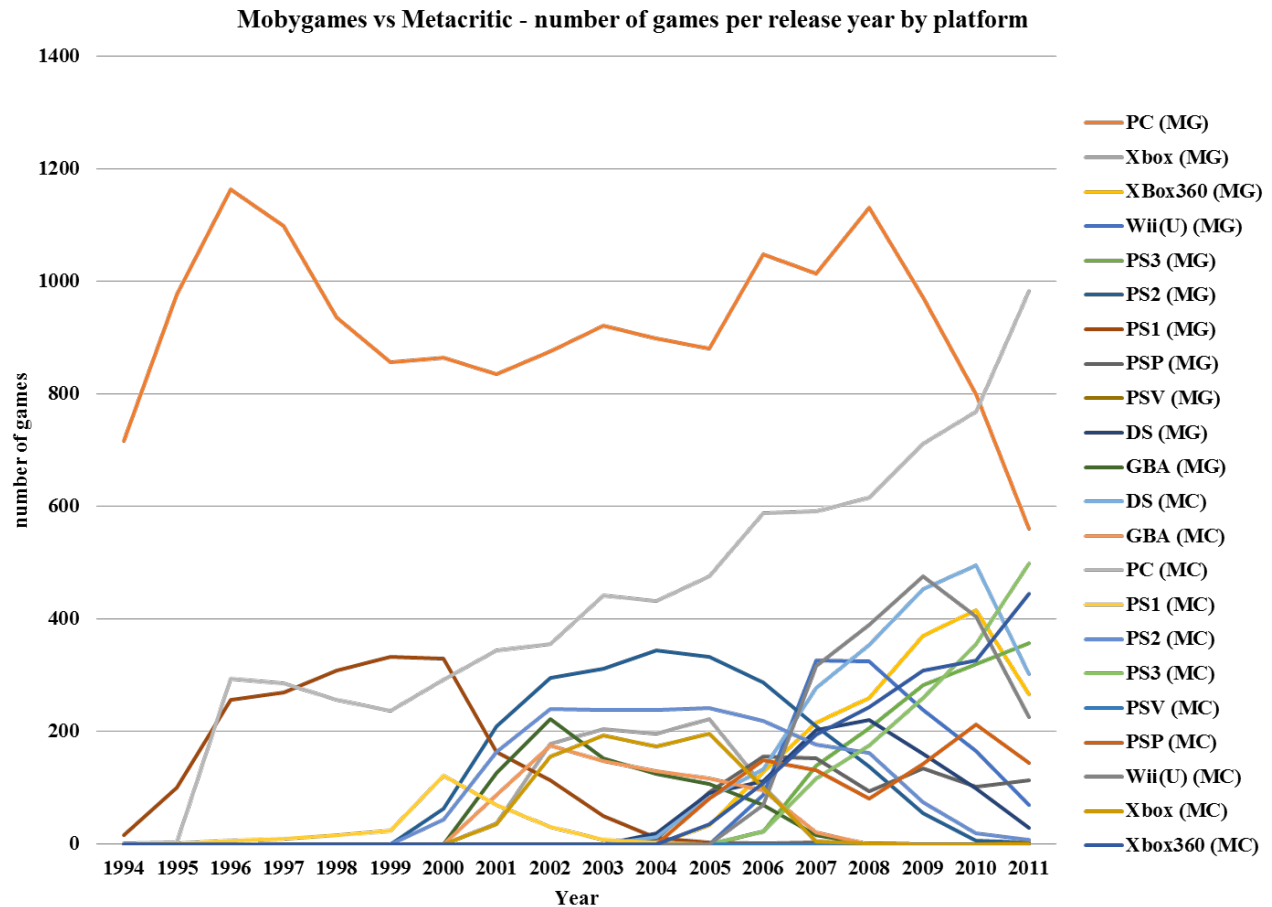


Figure A6. Comparison of the number of games listed at Mobygames (MG) versus Metacritic (MC) by year of release (1994-2011) and platform. Note that MG lists more PS1 and PC games in this period, though the curves for PC games cross at 2010, due to a steep decline in PC games listed at MG.