

<http://www.eludamos.org>

On the Validity of Metacritic in Assessing Game Value
Adams Greenwood-Ericksen, Scott R. Poorman, Roy Papp
Eludamos. Journal for Computer Game Culture. 2013; 7 (1), pp. 101-127

On the Validity of Metacritic in Assessing Game Value

ADAMS GREENWOOD-ERICKSEN, SCOTT R. POORMAN, AND ROY PAPP

In January 2001, the website Metacritic was launched with the goal of providing consumers with the ability to see a collection of game reviews in one location. The goal was admirable. Game reviews have long been scattered across a myriad of print and online media, and a consumer seeking several reviewer perspectives on the same game had to check multiple, unrelated information sources and then make judgments regarding the quality, accuracy, and content of each review in order to formulate an informed opinion on the quality of a product. Further, the scattered nature of such reviews meant that customers were often unable to easily identify which publications might have reviewed a game, making the process of determining which games to purchase an onerous chore. It appears that the founders of Metacritic hoped to change this paradigm by finding, indexing, and summarizing the scores provided by dozens of print and electronic media sources into a single, overall metascore. However, in recent years Metacritic has increasingly come under fire from critics who allege that it has become a harmful influence on the industry and that it fails to appropriately assess the value of individual games (Dodson 2006; Periera 2012; McDonald 2012). Therefore, the goal of the present work is to assess the scientific validity and empirical value of Metacritic as a tool to assess game value to both consumers and the industry.

The Origins of Metareview

The theory and practice of meta analysis was originally developed by scientists over a century ago (the first meta analysis is commonly attributed to the mathematician Karl Pearson and was conducted in 1904). The value of a meta analysis is twofold. First, it is able to aggregate many studies together statistically, allowing for a succinct and coherent analysis of the state of a body of research. Second, when many studies in an area show small effect sizes (the difference between possible outcomes is very small), meta analyses allow for stronger inferences to be made by looking at many less convincing studies together.

Scientific meta analyses are highly technical, in large part because of the complexity of the information being studied. However, by the close of the 20th century, a number of individuals and organizations on the web had noticed that a similar principle could be applied to the explosion of online and print reviews of popular consumer media. An early pioneer in this area was RottenTomatoes.com, which indexed, collected, and displayed movie reviews. The site opened on an amateur basis in 1999 (Lazarus 2001), and quickly became a popular source for movie information. Metacritic began in 2001 as an attempt by cofounders Marc Doyle, Julia Doyle Roberts, and Jason Deitz to extend the concept to a broader set of media (Wingfield 2007).

Why Metacritic Needs Assessment

The importance of Metacritic has grown significantly in recent years. Key figures in the game industry have made no secret of their concern with the scores assigned by Metacritic to games with which they have been involved). Interestingly, it appears that there is broad acceptance in the industry not only of the notion that Metacritic score impacts sales (Murdoch 2010; Wingfield 2007; Everiss 2008) but also that Metacritic is not a reliable assessment of game quality (Dodson 2006; Periera 2012; McDonald 2012). The present work is intended to address both of these issues through two related approaches. First, a correlational analysis of the relationship between Metacritic metascore and sales aimed at assessing the historical value of Metacritic scores as an indicator of financial game value will be presented. Subsequently, a comprehensive assessment of the scientific validity of the process by which Metacritic aggregates scores will be shown to demonstrate areas of logical methodological weakness in the Metascore production process. Taken together, these two analyses lead to the conclusion that while Metacritic is a strong predictor of sales, there are also significant flaws in the system by which Metascores are produced. The implications of these findings are also discussed.

Review of Literature

The stated goal of Metacritic is “helping consumers make an informed decision about how to spend their money on entertainment--by providing access to thousands of reviews in a number of entertainment genres” (Doyle 2011). Recently, however, Metacritic has come in for criticism from industry figures who argue that Metacritic is flawed and negatively impacting the health of the game industry (Dodson 2006; Periera 2012; McDonald 2012).

The Perception of Metacritic Score Impact on Sales

The internet is rife with opinions on the impact that metacritic has on game sales, many of them from apparent industry insiders. Regardless of the actual ground truth of the situation, the general perception of the relationship between sales and scores is worthy of discussion because of the impact that the opinions of decision makers can have on industry policies.

Overall, the general perception seems to strongly favor a clear link between sales and scores. John Riccitiello, CEO of Electronic Arts pointed out in a 2009 interview that “the best selling games in this industry last year were all 80 [Metacritic metascore] and above.” Julian Murdoch’s 2008 GamePro article “Metacritic: Gaming the Score” cites an interesting point made publicly by Robin Kaminsky, at the time VP of Marketing at Activision. During a presentation at DICE, a well regarded gaming business conference, Kaminsky declared that “for every additional five points over an 80 percent average review score, sales may as much as double” (Murdoch 2010). Similar sentiments have been attributed to Robert Kotik, CEO of Activision, who said “for every 5 percentage points [in metacritic score] above 80%, Activision found sales of a game roughly doubled” (Wingfield 2007; Everiss 2008). Peter Moore, a senior executive at EA Sports, initially espoused the use of Metacritic-based quality metrics,

but subsequently argued that they had become overused and might not be ideal development metrics (Dring 2010).

The Results of Perceived Sales Impact on Industry Policy

Due to this high level of acceptance of a direct relationship between sales and scores, it appears that at least some industry figures and studios have taken the apparently logical step of connecting scores to studio and employee valuation, and have implemented policies to support and incentivize high scoring games. After all, the argument goes, if scores equal sales, than scores equal value, and employees and studios should be incentivized to produce value by emphasizing the importance of metascores. John Riccitiello, EA's outspoken CEO, has commented publicly not only on the impact of the impact of metacritic scores on sales, but also on studio policy decisions, notably those related to compensation. "There are definitely bonuses attached to scores," he asserted in a 2009 interview that appeared on *Industrygamers.com*. Other sources have cited similar trends (Everiss 2008; Wingfield 2007).

There are certainly a number of possible implications of this trend. First, the impact on individual studios working with larger publishers can be significant. *Fallout: New Vegas*, the critically well-received fan favorite from Obsidian Entertainment was reportedly developed on contract to publisher Bethesda Softworks for a straight payment plus a bonus if the Metacritic metascore exceeded a value of 85. Unfortunately for Obsidian, the game apparently failed to meet that goal by one point, receiving a score of 84 (Gilbert, 2012). Interestingly, it appears that the original source for this information, a tweet from Obsidian veteran Chris Avellone has since been removed. As of March 15th 2012, it could be found at <https://twitter.com/#!/ChrisAvellone/status/180062439394643968>, but as of the time of this writing is no longer accessible at that address.

Metacritic scores may also have a broader impact on the external perception of viability or success for publishers or developers in the broader business community. For instance, THQ's *Homefront* received disappointing Metacritic metascores in the low to mid 70's across multiple platforms, apparently leading to upwards of a 20% drop in share price for the company (Pham and Fritz 2011; Baker 2011). The opposite effect has been observed as well, when Take-Two Interactive's stock price jumped 20% the week following the release of the critically-acclaimed *Bioshock* (Wingfield 2007). Note that despite widespread acceptance of claims to the contrary, there is no empirically valid way to connect these kinds of financial outcomes to metascores directly. Since metascores are based on widely-distributed reviews from independent critics and publications, it is just as reasonable to argue that the general response of critics (or potential purchasers themselves) or other factors such as seasonal buying patterns, marketing strategy, or word of mouth were responsible for the effect. Ultimately, however, these cases underscore the connection between perceived game quality and sales. Given that Metacritic is an aggregate indicator of critical response, however, it seems reasonable to suggest that metascore and sales might be connected. As of yet, however, there appears to have been no attempt to publish a broad analysis of the link between scores and sales, an oversight the present work aims to correct.

An interesting trend associated with this apparent relationship appears to be the tendency of some companies to develop design strategies explicitly aimed at maximizing metacritic score. Tim Heaton, studio director of Australia-based Creative Assembly (CA), has indicated in interviews that CA uses a strategy that specifically links features of games in production to actual hypothetical metascore points, and tracks expected metascore throughout the development process. The system apparently is used to estimate the impact of features with a very high degree of granularity, such that the impact of some features is apparently estimated down to at least the .5% metascore level (Nutt 2012).

Additionally, it appears that Metacritic metascores are also being used to determine hiring and compensation for individual employees. As discussed above, John Riccitiello, the CEO of Electronic Arts, has asserted this in the past (Brightman 2009), and similar claims have been advanced elsewhere (Everiss 2008; Wingfield 2007). Ultimately, it is probably fair to say that it is increasingly the case that individual developers may find that their compensation is directly tied to the metacritic scores of the games on which they work. This has been received in some quarters with hostility (Dodson 2006; McDonald 2012), but arguably represent a case of publishers and studios rewarding value with monetary compensation, assuming of course, that metascores are indeed a valid measurement of product value. Similarly, recently cases have emerged where metacritic scores have been explicitly linked to hiring decisions. On July 27, 2012, Irrational Games posted a job listing for a design manager which included the qualification requirement “credit on at least one game with an 85+ average Metacritic review score” (Graft et al. 2012). This reliance on Metacritic to drive hiring and compensation decisions for individuals raises further issues of fairness, especially in the context of the ongoing questions regarding the reliability of Metacritic metascores as an indicator of quality.

Unsurprisingly, this has resulted in a number of cases where developers or studios have resorted to tampering with Metacritic scores. In at least four documented cases, studio employees have been caught submitting user reviews for games they helped develop without acknowledging their studio affiliation (Sinclair 2011; Fahey 2011). It is unclear whether these individuals were acting with the knowledge of the leadership of the studios or publishers responsible for the games in question.

Ultimately, it seems reasonable to suggest that the perception of Metacritic throughout the game industry as an important metric of game quality has resulted in a broad swath of policies impacting everything from marketing strategy to the use of certain development approaches and metrics, and even to employee and studio compensation. As such, it seems that the influence of Metacritic on policies and decision-making in the game industry is both pervasive and powerful.

Criticism of Metacritic

Given the scope of the financial impact on all levels of the game industry associated with Metacritic metascores, it seems obvious that the fairness of this assessment system should be carefully examined. Certainly there is considerable criticism voiced among industry insiders at conferences and offices, although the authors have found this to be more true of off-the-record verbal communication than in written

publications. Some notable published criticisms do exist, of course. Joe Dodson's 2006 criticism of metareviews in general and Metacritic in particular deserves note (Dodson 2006). While hardly an unbiased (or even fair) criticism of metareview sites, the article did raise awareness of the controversy and made some reasonable points. It also may serve as a rough indicator of one branch of sentiment among game reviewers regarding metareviews.

It is clear that there are doubts about the validity of Metacritic as a source of unbiased feedback, even among those who promote its use. John Ricciatiello, for instance, who has been quoted previously in strong support of using Metacritic scores for various purposes, has also expressed reservations about its validity. "I'm a huge believer in quality, although I don't think Metacritic measures it the best for everything we do" (Brightman 2009). Peter Moore of EA Sports has been quoted as expressing reservations on the subject as well (Dring 2010). Given the increasing prevalence of Metacritic metascores as a primary indicator of game quality for both customers and industry decision-makers, and the financial implications thereof, it appears vital that a better understanding of the nature, origins, and validity of metacritic scores be undertaken.

Methods

The goal was to investigate whether a correlational link exists between game metascores obtained from Metacritic's website (<http://www.metacritic.com/>) and sales data as obtained from the website VGChartz (<http://www.vgchartz.com/>). These sources were chosen in part because they are readily accessible to members of the industry and the general public, which should make it easier for others to replicate and extend the current work independently.

Sampling

A random sample of 196 Games was drawn from Metacritic. Games were selected from the Action, RPG, and FPS genres, as defined by Metacritic's internal classification system. Only games released for the XBOX 360 and Playstation 3 consoles were chosen because of the relative similarity of marketing, deliver, and control systems between titles released for the two platforms. A listing of the games included in the sample, as well as associated sales and score data is included in appendix A, below. Sales data were then obtained (in millions of units) from the website VGChartz. Metacritic score and sales data were collected in August of 2010, and reflect the information available from those sources at that time.

Analysis Approach

The data collected were analyzed in a three-step process. First, they were plotted out on a graph to allow visual identification of patterns and characteristics of the data. Then, a statistical measure known as a "Pearson's r," or "Pearson Product-Moment

Correlation Coefficient” was applied to the data to identify the correlation between the two data sets.

Visual Analysis

A graphed plot of all of the data was performed in order to visually identify broad patterns in the data. Plots were performed for the entire data set ($N = 196$), as well as for each individual combination of platform (PS3, XBOX360) and genre (Action, RPG, and FPS). Visual plots of the data are presented in the below, grouped by genre (Figure 1) and by platform (Figure 2). Visual inspection of the data appeared to show a meaningful geometric or exponential relationship between sales and scores.

Quantitative Analysis

The collected data were subsequently analyzed using Pearson’s Product-Moment Correlational Coefficient (PMCC, or Pearson’s r). Because the visual analysis indicated a pronounced curve to the data set, analysis was split into two parts: first, a bivariate correlation using the untransformed data set was used to assess the linear relationship discounting the obvious visible curve. Such an analysis involves the least amount of processing of the data and therefore might be seen as a more conservative statistical analysis approach. However, such an approach would be expected to underestimate the relationship between the variables, and furthermore violates the assumption of linearity inherent in the PMCC. Therefore, a second analysis was completed after applying a logarithmic transformation to both variables to “flatten out” the curve of the data. This approach, although it involves more processing of the data, should be expected to yield a more accurate coefficient of correlation. Both analyses are presented so that the reader can judge for themselves which they prefer. Note that these two reported analyses should be seen as alternative approaches, rather than one confirming or reinforcing the findings of the other.

Results

Bivariate Correlational Analysis

A Pearson’s product-moment correlation coefficient (PMCC) was calculated on the untransformed data set and showed a significant positive correlational relationship between sales and scores, $r = .55$, $p < .005$. Of course, given the apparent curvature of the data, it is expected that the relationship between sales and scores might be seriously underestimated by this procedure, given that the PMCC assumes a linear relationship between data sets. However, it was expected that the results of this rather unsophisticated analysis approach using untransformed data would nonetheless show a meaningful relationship and would help alleviate any concerns about the conservativeness of subsequent transformation-based analyses.

Transformation of Data

Because the data plot suggests a nonlinear relationship between metascore and sales, the above analysis on untransformed data almost certainly underestimates the strength of the relationship between the two variables, as linearity is an assumption of the PMCC. Although the analysis on the untransformed data set still shows a significant correlation, in the interests of fully understanding the relationship between the variables, a more satisfying and accurate approach can be achieved by transforming the data to achieve linearity before calculating the bivariate correlation. In this case, a log transformation was chosen because of its efficacy in linearizing curvilinear data sets. The transformed results also suggested a significant positive relationship between sales and scores, $r = .72$, $p < .005$. The increase in the reported r value for the PMCC indicates an even stronger relationship between sales and scores than that suggested by the analysis on untransformed data.

Analysis Summary

The results of our analyses are shown below. Visual analysis of the graph shows an apparent geometric or exponential relationship between game sales and metascore, such that higher metascores are associated with higher sales. Additionally, the curve appears to have a “break point” somewhere around 80% where the rate of increase in sales begins to trend strongly upwards.

The initial correlational analysis showed a correlation of .55 on a scale of -1 to 1, which is generally considered to be a reasonably large correlation. The correlation was statistically significant at the .005 level (a criterion ten times more stringent than is typical for these analyses). However, because correlational statistics are designed for linear data rather than curvilinear data, we also performed a second analysis after applying a mathematical procedure known as a “log transformation” to “straighten” the data set. A correlational analysis of the transformed data set revealed a new correlation of .72, far higher than even the initial estimate. This result was also statistically significant at the .005 level, indicating a very high level of confidence in the result.

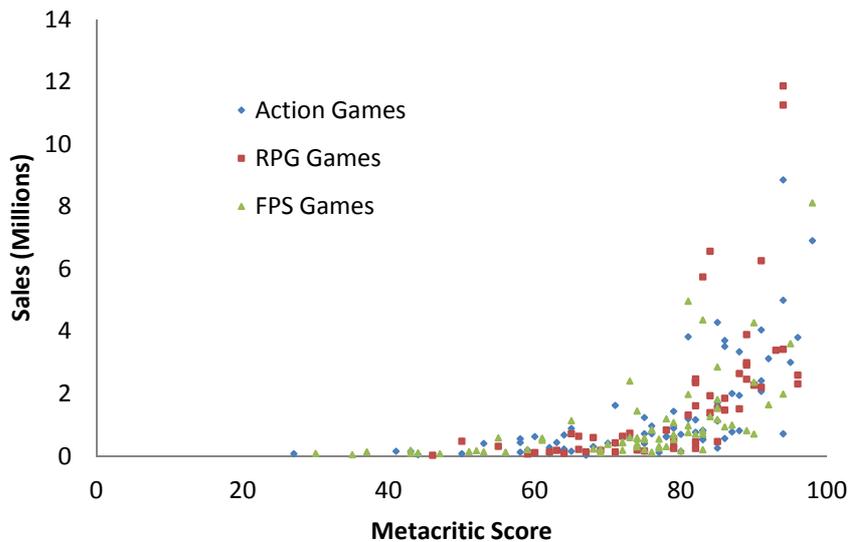


Figure 1. Metacritic Score versus Sales (in Millions) by Genre

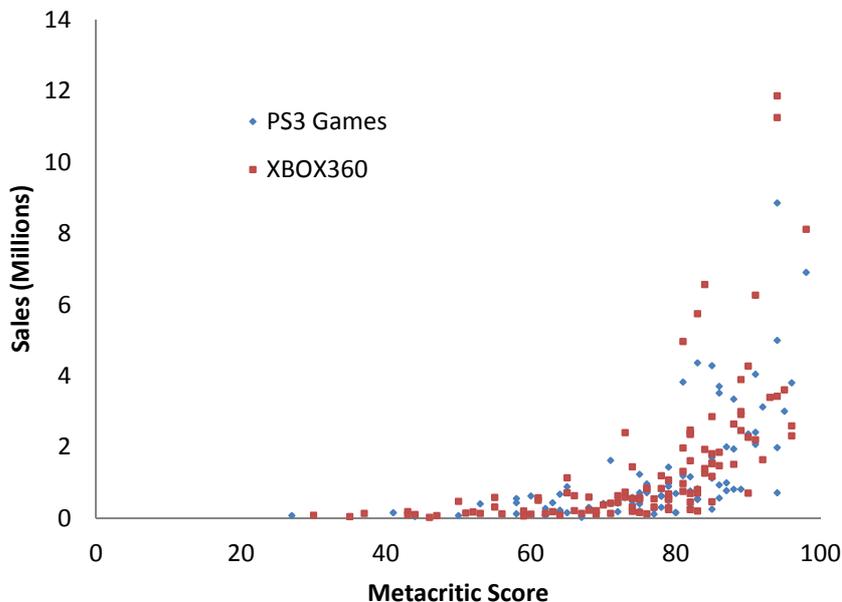


Figure 2. Metacritic Score versus Sales (in Millions) by Platform

Discussion

The dual approach used in the present work was intended to examine the issues surrounding Metacritic scores from both a qualitative and quantitative perspective. The quantitative examination of the mathematical relationship between sales and scores using publicly available data was intended to address the issue from an empirical and number-driven perspective. The tight coupling between sales and scores strongly suggests that Metacritic is a valuable tool for assessing (and possibly predicting) game value in terms of critical acclaim, sales, and return on investment for studios and publishers. While the quantitative analysis above has provided strong evidence of a significant relationship between sales and scores, such an approach

cannot shed light on the validity or reliability of the procedures by which Metacritic calculates metascores. To address these concerns, a qualitative analysis of the metascore generation processes was conducted. By carefully examining the validity issues with metacritic from a scientific perspective, it was hoped that insights could be gained into how score validity and reviewer intent was preserved or distorted at each step in the process, as well as how this process would impact the overall value of Metacritic as a tool for decision-makers in the game industry.

Qualitative Analysis of Validity

Scientists typically discuss the quality of a measure or argument in terms of causal “validity,” or simply “validity.” Scientists generally recognize five subcategories to validity, each of which pertains to a specific aspect of the measurement or argument in question. Because Metacritic is essentially drawing a conclusion about the quality of a game based on a rating developed using a mathematical argument (Metacritic’s proprietary formula) which incorporates a number of measured data points (individual scores), it is vulnerable to concerns about the validity of the process used to make these determinations. Since not everyone is a scientist, a discussion of causal validity as it applies to an assessment of Metacritic is included below.

Internal Validity is about whether a measurement is being assessed in such a way as to determine the appropriate cause for a given effect. An example of this is the classic chicken-egg problem: do chickens cause eggs, or do eggs cause chickens? In the case of metacritic, key questions include, for instance, which review sites are being polled, whether external events have an impact on individual reviews (other reviews, reviewer-developer relationships, etc), and other, similar concerns.

Construct Validity is about whether a documented scale is measuring what it is supposed to be measuring, or something else entirely. IQ tests, for instance, are notorious for measuring things other than intelligence (educational background or ethnicity, for instance). In the case of Metacritic, one interesting question is how the different scales used by different reviewers and publications are “normalized” to fit Metacritic’s 100 point scale, and whether distortion of the reviewer’s intent occurs during the process.

External Validity focuses on whether a measurement or finding is likely to generalize outside of the specific conditions where the test occurred. In the case of Metacritic, one key question is whether reviewers are a good approximation of customers with regards to the things they like and dislike. Another, partially addressed above, is whether metascore correlates with other real world measures of game success, such as sales or awards.

Face Validity is an indicator of how good a measurement or argument *appears* to be. This is similar to Stephen Colbert’s concept of “truthiness” (which as of 2011 appeared in the Oxford English Dictionary). Just as an idea that is “truthy” appears to be or “feels like” the truth, whether it is actually true or not, a measurement or argument that shows good “face validity” seems like it should be right, regardless of whether or not it actually is. Metacritic typically enjoys high face validity in many

circles, as it appears (on the surface at least) to be an unbiased aggregate overall score.

Statistical Conclusion Validity assesses whether the mathematical or statistical procedures used on the data are appropriate. This can be highly technical in the case of complicated experiments, but in the context of Metacritic this mostly boils down to whether Metacritic's approach to the mathematical aggregation of game review information could reasonably be expected to yield an accurate representation or assessment of game quality.

How Metacritic metascores are calculated

Marc Doyle and other Metacritic employees have been reasonably forthright on the subject of exactly how Metacritic calculates metascores. The website itself presents a layman's description of the process: "We carefully curate a large group of the world's most respected critics, assign scores to their reviews, and apply a weighted average to summarize the range of their opinions" (Metacritic 2012a). The site goes on to explain that:

Metascore is a weighted average in that we assign more importance, or weight, to some critics and publications than others, based on their quality and overall stature. (Metacritic 2012a)

This is an important point, as it illustrates one of the aspects of this process that often attracts the strongest criticism and confusion. By applying a mathematical "weight" to each individual score, Metacritic is asserting that the opinions of some publications or critics are more important than others. Predictably, this is not received well in all circles (Dodson 2006). Regardless, it appears that Metacritic follows the steps illustrated in Table 1 below when preparing and delivering a metacritic score.

Step	Action taken by Metacritic
1	Identify "trusted" publications and critics from which it will draw scores.
2	Assign a "weight" to each of these based on how much Metacritic trusts or respects their work and judgment.
3	Gather individual reviews from these publications and critics
4	Apply Metacritic's conversion scales to the original publication score
5	Aggregate all scores into a weighted average using the individual scores from step 3 and the weights from step 2.
6	Publish these metascores on their website at metacritic.com

Table 1: Steps in Metacritic's metascore creation process

The ultimate outcome of this process is a single measurement that incorporates not only the individual score contributed by the critic or publication, but also Metacritic's assessment of the worth or reliability of that source.

The validity of Metacritic metascores

As with any process related to subjective criticism, there are a number of areas of concern with regards to the calculation of metacritic scores. Table 2 below summarizes some relevant concerns at each step of the broader assessment process (including the contribution of the original critic or publication).

Action	Associated Potential Threats to Validity
Individual reviewer assigns a score based on their own opinion and scoring system.	<p>Reviewer can be biased for or against the game, genre, series, studio, or publisher for any of a number of reasons.</p> <p>Reviewer can be influenced by previous iterations in a series.</p> <p>Reviewer can be influenced by other published scores for the game in question.</p>
Metacritic gathers scores from individual sites	<p>Metacritic may miss a score from a publication or critic that they intend to track</p> <p>Some important or useful scores may not be considered because Metacritic does not track them.</p> <p>Metacritic staff may misinterpret a reviewer's intent when assigning a score to reviews in which no quantitative score is provided.</p>
Metacritic applies conversions to 100 point scale	Metacritic's conversion system may distort the reviewer's intent (see Tables 3, 4, and 5 below).
Metacritic aggregates all scores into a weighted average	<p>Weighting may not accurately represent the general consensus of reviewers.</p> <p>Weights are assigned at the discretion of Metacritic and criteria for weighting are not transparent.</p> <p>A single highly divergent score from a highly-weighted publication can distort the overall metascore.</p>
Metacritic publishes the metascore	Consumers can misunderstand the meaning, relevance, or importance of a metascore.

Table 2: Potential threats to validity associated with Metacritic's metareview process

The first, and in some ways the most basic, potential problem with Metacritic metascores is the inherently subjective nature of critical review. Not all critics agree on the quality of a given art object, product, or service (as games could potentially be categorized as any of these, depending on features and/or distribution approach). An examination of the possibilities for a breakdown during critical review is well beyond the scope of this work, and isn't entirely germane to the issue of *Metacritic's* validity specifically, but is still important to note. At a minimum, there are several types of issues related to critical review as a basic level that need to be considered:

1. Issues of reviewer bias stemming from reviewer attitudes toward the game, publisher, genre, development studio, or content area.
2. Reviewer experience with game genres, games, or criticism in general.
3. Editorial pressure stemming from personal or financial relationships between publishers or studios and publications.
4. Reviewer peer pressure stemming from previously published reviews of the game in question.

All of these issues should be matters of concern when considering the reliability and accuracy of game reviews, and these represent fertile topics for future research. However, the focus of the present work is on the impact that *Metacritic itself* as an organization or information source has on the process.

Gathering Individual Reviews

Even if all reviews are reasonably on-target, a number of other potential pitfalls emerge as these reviews make their way into Metacritic's database. First, Metacritic makes it clear that they do not track reviews from all publishers. The actual requirements for inclusion in Metacritic are not entirely transparent, but appear to include publication reputation, subjectively-assessed review quality, and review quantity (Metacritic, 2010c). Therefore, it is entirely possible that a review for a given game may appear in an untracked publication or source, and would therefore not be included in Metacritic's score. Further, although representatives of Metacritic have previously stated that there are certain publications that are regularly checked for reviews (Metacritic, 2012b), Metacritic's staff may not become aware of a particular review of a game that appears in a tracked publication, either as a result of an oversight or because the review is not noticed by or is inaccessible to their staff. Therefore, many relevant reviews may be overlooked either because the publication in question is not tracked, or because of a failure in the review collection or tracking process.

Even when a review is identified, there are certain cases where reviewers do not assign a score to a game. Under those circumstances, it is Metacritic's policy to also assign a score to a review when none exists based on a subjective assessment of reviewer intent by Metacritic staff (Metacritic 2010a). Given the inherently inconsistent nature of subjective assessment, and the lack of inside knowledge of the reviewer's state of mind on the part of Metacritic staff, it is obviously possible that the reviewer's intent may not be appropriately understood and documented, posing a serious threat to validity.

Score Conversion

Different reviewers and publications use widely varying methods for quantifying the quality of a game. In cases where a quantitative score is available directly from the reviewer, Metacritic generally needs to convert the score used by the publication or reviewer into Metacritic's 0-100 scale format in order for it to be included in their database. Metacritic clearly lists their conversion system on their website with tables for 4-star scales (Table 4), traditional A-F scholastic grading scales (Table 5), and the rather obvious 1-10 scale conversion (Table 3). The conversion systems for other scales (thumbs up/thumbs down, go/wait/don't go, buy/rent/skip, etc) are not included on Metacritic's website, and do not appear to be published elsewhere. It seems likely that these represent cases where Metacritic staff subjectively assign a 0-100 score directly, consistent with their policy as documented in Metacritic (2010a).

The translation of scores from one scale to another is a very problematic process from a validity perspective. Many of the difficulties lie in the distinction between the perceptions of the reviewer, the general public, and Metacritic staff on the meaning of certain specific ratings, particularly when there are preexisting problems with the scale used in the review.

This is nowhere seen more clearly than in the A-F conversion scale used by Metacritic, although it can be argued that the difficulty in this case is not really of Metacritic's making. The traditional A-F scholastic grading scale has long been known to be quite seriously flawed, most obviously with regards to a problem known as "restriction of range." Typically, a score of 100-90 is seen as an "A," an 89-80 as a "B," a 79-70 as a "C," and so on. Some scales use "+" and "-" modifiers to increase the granularity of the score, such that 100-97 is seen as an A+, a 96-94 as an A, a 92-90 as an A-, a 89-87 as a B+, an 86-84 as a B, and so on. Regardless of which type of A-F scale is used, it quickly becomes apparent that the lowest possible grade (an "F"), includes the entire set of scores from 50-0, making the range covered by "F" anywhere from 5-15 times larger (depending on whether or not one includes "+" and "-" grades) than any other category.

While broad exposure has conditioned individuals in the United States and other countries which commonly use this scale to accept the qualitative value of each of these grade categories, translating scores from varying rating systems into a true 0-100 scale represents a serious challenge, because the A-F system is actually based on only half of the 100 point range, as everything at or below 50 is simply an "F." This puts Metacritic in the unenviable position of choosing between using only half of their overall scale (thereby potentially artificially inflating the score above what was intended by the reviewer), or having to redefine the numeric values associated with each letter grade contrary to the established public perception of their value. By choosing the latter approach, Metacritic faces the validity challenge of a serious discrepancy between the numbers many reviewers *expect* will be associated with a letter grade, and those that are actually applied by Metacritic.

For instance, it can be seen in Table 5 that Metacritic assigns a score of 75 to a game rated as a B, and 67 to a game rated as a B-. This is of course confusing to individuals who consider a "B" in the context of the A-F scholastic grading scale to be a reasonably good outcome (typically, an 86-84%). By contrast, a 75 is seen as a weak grade. This discrepancy is even more pronounced for games with B-, C, D, and

F ratings. The result is that Metacritic's conversion system may distort either the perception of the user as to what a score means, the intent of the reviewer, or both. This particular discrepancy has been widely documented elsewhere (Wingfield 2007; Boesky 2008).

Other scales have their own conversion problems. On a 4-star scale, removing a single star drops a game to a 75% rating once the conversion is applied (see table 4). This low level of granularity may result in the artificial deflating of a score contrary to the intent of the reviewer, and may additionally cause confusion among user of Metacritic.

Publication's rating	Metacritic's Rating
10	100
9	90
8	80
7	70
6	60
5	50
4	40
3	30
2	20
1	10
0	0

Table 3: Metacritic score conversion: 10 to 100 point scale (from Metacritic 2012a)

Publication's rating	Metacritic's Rating
4 stars	100
3.5 stars	88
3 stars	75
2.5 stars	63
2 stars	50
1.5 stars	38
1 stars	25
0.5 stars	12
0 stars	0

Table 4: Metacritic score conversion: X out of 4 Stars to 100 point scale (from Metacritic, 2012a)

Publication's Rating	Metacritic's Rating
A or A+	100
A-	91
B+	83
B	75
B-	67
C+	58
C	50
C-	42
D+	33
D+	25
D-	16
F+	8
F or F-	0

Table 5: Metacritic score conversion: A-F to 100 point scale (from Metacritic, 2012a)

Score Aggregation and Weighting

Metacritic has been quite open and consistent in stating that their metascores are calculated using a weighted average (Metacritic 2012a), which is calculated by multiplying each score by a coefficient that is used to represent the quality or importance of the individual score in assessing the game as a whole. However, Metacritic has previously refused to comment on the specific weights they apply to various publications or reviewers in calculating the value of metacritic scores (Metacritic 2010b). This is understandable for several reasons. First, this represents some level of proprietary system for Metacritic, and could be seen as a form of intellectual property. Second, it represents a potentially volatile issue with regards to the public perception of certain reviewers and publications. The nature of this type of rating system inherently implies value judgments about the quality of publications and reviewers, which of course makes this information highly sensitive and potentially controversial in nature. Additionally, little information is available at the present date regarding the exact process by which Metacritic assigns and maintains these weights.

Score Presentation

Even once this process is complete, there remains one remaining potential problem – the color code used by Metacritic based on the numeric metascore a game receives. Scores in the 100-75 range are displayed to the user in green colored text, scores from 74 to 50 in yellow, and 49 and below in red (see table 6 below for detailed breakdowns as published by Metacritic). This may be seen as implying a more judgmental assessment on the part of metacritic, such that green games are “good,”

yellow games are “moderate” in quality, and red games are “bad.” Additionally, the sharpness of the rating scale means that a game that scores a 74 and thereby misses the green color category by a single point (1/100th) of the scale gets the same color code as a game that gets a 50. Taken together, these problems could lead to the distortion of user perceptions such that users of Metacritic’s site may perceive certain games as being much better than others when the actual difference is much more subtle.

General Meaning of Score	Games	Color
Universal Acclaim	90 - 100	Green
Generally Favorable Reviews	75 - 89	Green
Mixed or Average Reviews	50 - 74	Yellow
Generally Unfavorable Reviews	20 - 49	Red
Overwhelming Dislike	0 - 19	Red

Table 6: Metacritic score conversion: 100 point scale to color code (from Metacritic 2012a)

Qualitative Validity Assessment

The investigation of validity in the first part of the paper identified a number of flaws in the methodology used by Metacritic to calculate metascores. Many of these deal explicitly with the translation of the intent of the reviewer to a 0-100 numeric score, but threats to validity and accuracy have been identified at every step of the process. Taken together, these findings raise concerns regarding the accuracy and validity of metascores as representations of the aggregate opinion of the community of game reviewers regarding the quality and value of specific games.

The issues associated with the translation of various reviewer scales to Metacritic’s 100 point scale are particularly worrisome, not only because they affect the actual reliability of metacritic as an assessment tool with regards to how appropriate the assessment process is (internal validity) and how reliable the measurements are (construct validity) but also because they seem to have a broad impact on the perception of the reliability of Metacritic as a whole (face validity).

However, these analyses are only half the story – equally important is direct observation of the actual accuracy and consistency with which Metacritic metascores predict or correlate with actual game sales.

Empirical Results

In general, the results of the statistical analysis in the results section above showed a very strong relationship between sales and scores, regardless of genre or platform. This is a very important point, as much of the criticism of Metacritic metascores centers around the idea that they fail to accurately represent product value. Our

results showed fairly conclusively that there was a tight coupling between sales and scores, such that games with higher Metacritic metascores tended to have higher sales as well, across genre and platform. Accordingly, despite the threats to validity noted in earlier sections, it is difficult to argue against the value of Metacritic as an assessment tool when it shows itself to be such a clear bellwether of financial success in games.

Some important caveats exist, however. First, by the nature of the mechanisms of assessment available to the authors, the data collected are necessarily observational in nature; that is to say, that they allow us to talk about the *correlation* between sales and scores, but do not allow us to say definitively whether high scores *cause* high sales, or the converse, or whether a more complicated relationship exists involving other factors such as marketing or media exposure. One obvious interpretation would be that both high scores and high sales are correlated with game quality, which while gratifying to proponents of Metacritic, is unfortunately only one of many possible explanations. Ultimately, the most likely interpretations would appear to be that (1) Metacritic is driving sales, (2) Metacritic is predicting sales, that (3) both Metacritic score and sales are both being driven by a third factor such as game quality, reviewer bias, or marketing activity, or that (4) some combination of the above factors is in play. Regardless, the important point is that the strong relationship between the two would seem to suggest that Metacritic is a good benchmark for studios and publishers interested in assessing the financial value of individual games, whatever the industry or general public may think of its suitability as a measure of game quality.

Correcting the Flaws in Metacritic

Despite the identification of serious concerns regarding the process Metacritic uses in gathering and aggregating scores, it is difficult in many cases to see how Metacritic could act to address them. Many of these, such as the score translation problem, arise either from the inherent drawbacks of metareviews in general, or as a result of decisions made by individual reviewers whose choices are outside of the control of Metacritic as an organization. For instance, an A-F rating scale is broadly regarded as a flawed scale, replete with validity issues on multiple levels, yet it continues to be used by some publications and reviewers (as well as most school districts in the United States). No action on the part of Metacritic, other than entirely excluding any score not formatted as a 0-100 scale, would address such scale translation issues entirely. Further, were Metacritic to take that drastic step, it would arguably produce a far worse outcome by providing a score based only on a few specific sources of reviews and excluding a large number of valid perspectives on a given game title.

One method of addressing criticisms of the “one size fits all” model of metascore generation (i.e. that it assumes that all users have the same tastes) might be the adoption of a more sophisticated individualized approach, in which users are provided with relative ratings for games based on their stated preferences or user review history. This could have the dual benefit of defusing the “absolute measurement” value that has attracted so much negative attention to Metacritic while providing a more personally relevant score to each particular user. The potential

improvement in industry acceptance and specific user-focus might well be worth the increased complexity inherent in implementing such an approach.

Additionally, adding additional transparency to the weights and formula Metacritic uses to calculate metascores could help to reduce the mystery of how scores are calculated (and could thereby reduce suspicion on the part of industry members and users).

Conclusions

Overall, debate on this issue will almost certainly continue. However, a few things can be said with a fair degree of confidence. First, Metacritic's process for gathering, translating, and aggregating scores appears to be flawed at several levels. That being said, in many cases it is unclear how precisely these flaws could or should be addressed. Other issues may well be systemic to the community of game reviewers and publications. This factor may be particularly problematic to address because these individuals and groups do not appear to adhere in many cases to basic standards of journalistic and editorial professionalism. Examples of such standards which are routinely neglected by industry-targeted publications include avoiding or disclosing conflicts of interest on the part of reviewers and publications, clearly differentiating between paid or advertising content and news or opinion material, and consistently requiring relevant educational credentials or certifications of reviewers.

Ultimately, it is difficult to escape the conclusion that the strong empirical evidence for a close link between sales and scores argues strongly for the value of Metacritic as an assessment tool. Accordingly, it is naïve to expect publishers or other decision makers in the industry to abandon Metacritic as a yardstick anytime in the foreseeable near-term future. Indeed one might expect them to adopt the tool more fully in that role. The cases of *Homefront* and *Bioshock* also clearly indicate that financial markets and the broader business community consider Metacritic to be an important indicator of product quality and therefore company health, and are likely to continue to make judgments of the value of games based on metascores. This cannot but help have the effect of further raising the profile and importance of Metacritic scores even higher among shareholders, executives, and the general public. Additionally, the financial success of Metacritic and its high visibility indicate that it has come to play a significant, if not central, role in driving consumer purchasing decisions. Future research could certainly be done to productively establish precisely the nature of that relationship, but in the meantime the industry should probably expect the influence of Metacritic to increase, rather than decrease.

One additional note of caution is in order as well – it may well be the case that since Metacritic acknowledges that their metascore formula is based on a weighted average, the only intellectual property of value that the company possesses, aside from its current visibility, is the proprietary list of reviewer weightings they use to derive these scores. The simplicity of Metacritic's approach to calculating aggregated metareviews may therefore make it potentially vulnerable to upstart competitors who utilize more sophisticated approaches to calculate or display and visualize data. If someone else finds a better, more easily accessible way to do what

Metacritic currently does, the organization could quickly experience a ruinous fall from their current ascendancy.

Appendices

Appendix A: Scores vs. Sales Data				
Genre	Game	System	Score	Sales (in Millions)
Action	Naughty Bear	XBOX 360	43	0.19
Action	The Lord of the Rings: Conquest	XBOX 360	55	0.59
Action	Dark Void	XBOX 360	59	0.21
Action	Avatar: The Game	XBOX 360	61	0.58
Action	Transformers: Revenge of the Fallen	XBOX 360	61	0.51
Action	Ninja Blade	XBOX 360	68	0.23
Action	Deadly Premonition	XBOX 360	69	0.12
Action	Silent Hill: Homecoming	XBOX 360	70	0.38
Action	Dante's Inferno	XBOX 360	73	0.6
Action	Star wars: The force Unleashed	XBOX 360	73	2.41
Action	Mafia II	XBOX 360	74	0.56
Action	Prince of Persia: The Forgotten sands	XBOX 360	74	0.3
Action	X-Men Origins: Wolverine	XBOX 360	75	0.58
Action	Prototype	XBOX 360	78	1.2
Action	Dead Rising 2	XBOX 360	79	0.69
Action	Ghostbusters: The Video game	XBOX 360	79	0.54
Action	Mirror's Edge	XBOX 360	79	1.08
Action	Assassins Creed	XBOX 360	81	4.97
Action	Just Cause 2	XBOX 360	81	0.76
Action	Ninja Gaiden II	XBOX 360	81	0.98
Action	Saints Row 2	XBOX 360	81	1.98
Action	Brutal Legend	XBOX 360	82	0.7
Action	Alan Wake	XBOX 360	83	0.81
Action	Castlevania: Lords of Shadow	XBOX 360	83	0.21
Action	Darksiders	XBOX 360	83	0.71
Action	Devil May Cry 4	XBOX 360	84	1.27
Action	Dead Rising	XBOX 360	85	1.82

Action	Resident Evil 5	XBOX 360	85	2.86
Action	Tom Clancy's Splinter Cell: Conviction	XBOX 360	85	1.55
Action	Tom Clancy's Splinter Cell: Double Agent	XBOX 360	85	1.18
Action	Assassins Creed II	XBOX 360	90	4.28
Action	Bayonetta	XBOX 360	90	0.71
Action	Batman: Arkham Asylum	XBOX 360	92	1.65
Action	Red Dead Redemption	XBOX 360	95	3.61
Action	Grand Theft Auto IV	XBOX 360	98	8.12
Action	Iron Man 2	PS3	41	0.16
Action	Naughty Bear	PS3	43	0.16
Action	Lair	PS3	53	0.41
Action	Fist of the North Star: Ken's Rage	PS3	58	0.56
Action	Way of the Samurai 3	PS3	58	0.44
Action	Dark Void	PS3	59	0.2
Action	Avatar: The Game	PS3	60	0.63
Action	Dynasty Warriors 6 Empires	PS3	62	0.28
Action	Transformers: Revenge of the Fallen	PS3	63	0.44
Action	Silent Hill: Homecoming	PS3	64	0.23
Action	Star wars: The force Unleashed	PS3	71	1.63
Action	X-Men Origins: Wolverine	PS3	73	0.59
Action	Dante's Inferno	PS3	75	0.72
Action	Prince of Persia: The Forgotten sands	PS3	75	0.41
Action	Spider-Man: Shattered Dimensions	PS3	75	0.17
Action	Ghostbusters: The Video game	PS3	78	0.62
Action	Heavenly Sword	PS3	79	1.44
Action	Mirror's Edge	PS3	79	0.9
Action	Prototype	PS3	79	0.98
Action	Assassins Creed	PS3	81	3.83
Action	Darksiders	PS3	82	0.77
Action	Saints Row 2	PS3	82	1.17
Action	Brutal Legend	PS3	83	0.55
Action	Just Cause 2	PS3	83	0.83
Action	Ninja Gaiden Sigma 2	PS3	83	0.53
Action	Devil May Cry 4	PS3	84	1.31

Action	Castlevania: Lords of Shadow	PS3	85	0.26
Action	Infamous	PS3	85	1.71
Action	Resident Evil 5	PS3	86	3.52
Action	Bayonetta	PS3	87	0.78
Action	Uncharted: Drake's Fortune	PS3	88	3.35
Action	Assassins Creed II	PS3	91	4.05
Action	Batman: Arkham Asylum	PS3	91	2.08
Action	God of War III	PS3	92	3.13
Action	Metal Gear Solid 4: Guns of the Patriots	PS3	94	5
Action	Red Dead Redemption	PS3	95	3.01
Action	Uncharted 2: Among Thieves	PS3	96	3.81
Action	Grand Theft Auto IV	PS3	98	6.91
FPS	ShellShock 2: Blood Trails	XBOX 360	30	0.09
FPS	History Channel: Battle for the Pacific	XBOX 360	35	0.05
FPS	Hour of Victory	XBOX 360	37	0.14
FPS	America's Army: True Soldiers	XBOX 360	43	0.1
FPS	NPPL Championship Paintball 2009	XBOX 360	44	0.11
FPS	Legendary	XBOX 360	47	0.08
FPS	History Civil War: Secret Missions	XBOX 360	51	0.15
FPS	Conflict: Denied Ops	XBOX 360	52	0.18
FPS	History Channel: Civil War- A Nation Divided	XBOX 360	53	0.14
FPS	Velvet Assassin	XBOX 360	56	0.13
FPS	007: Quantum of Solace	XBOX 360	65	1.14
FPS	Section 8	XBOX 360	69	0.22
FPS	Wolfenstein	XBOX 360	72	0.45
FPS	Medal of Honor	XBOX 360	74	1.45
FPS	Frontlines: Fuel of War	XBOX 360	75	0.54
FPS	Brothers in Arms: Hell's Highway	XBOX 360	76	0.84
FPS	Operation Flashpoint: Dragon Rising	XBOX 360	76	0.85
FPS	Singularity	XBOX 360	76	0.13
FPS	Call of Juarez: Bound in Blood	XBOX 360	77	0.55
FPS	Metro 2033	XBOX 360	77	0.32
FPS	Prey	XBOX 360	79	0.3
FPS	Perfect Dark Zero	XBOX 360	81	1.32

FPS	Call of Duty 3	XBOX 360	82	2.36
FPS	The Chronicles of Riddick: Assault on Dark Athena	XBOX 360	82	0.25
FPS	Tom Clancy's Rainbow Six: Vegas 2	XBOX 360	82	2.39
FPS	Unreal Tournament III	XBOX 360	82	0.46
FPS	Halo 3:ODST	XBOX 360	83	5.75
FPS	Battlefield: Bad Company	XBOX 360	84	1.4
FPS	Borderlands	XBOX 360	84	1.94
FPS	Call of Duty: World at War	XBOX 360	84	6.57
FPS	Far Cry 2	XBOX 360	85	1.54
FPS	FEAR 2: Project Origin	XBOX 360	85	0.47
FPS	Tom Clancy's Ghost Recon Advanced Warfighter 2	XBOX 360	86	1.48
FPS	Battlefield: Bad Company 2	XBOX 360	88	2.65
FPS	Bioshock 2	XBOX 360	88	1.52
FPS	Call of Duty 2	XBOX 360	89	2.47
FPS	Left for Dead	XBOX 360	89	2.92
FPS	Left for Dead 2	XBOX 360	89	3
FPS	Tom Clancy's Ghost Recon Advanced Warfighter	XBOX 360	90	2.28
FPS	Halo Reach	XBOX 360	91	6.27
FPS	Call of Duty: Modern Warfare 2	XBOX 360	94	11.87
FPS	Halo 3	XBOX 360	94	11.26
FPS	Bioshock	XBOX 360	96	2.6
FPS	Rogue Warrior	PS3	27	0.08
FPS	Soldier of Fortune: Payback	PS3	50	0.08
FPS	007: Quantum of Solace	PS3	65	0.89
FPS	Wolfenstein	PS3	71	0.42
FPS	Medal of Honor	PS3	75	1.24
FPS	Brothers in Arms: Hell's Highway	PS3	76	0.73
FPS	MAG	PS3	76	0.97
FPS	Operation Flashpoint: Dragon Rising	PS3	76	0.73
FPS	Singularity	PS3	77	0.12
FPS	Call of Juarez: Bound in Blood	PS3	78	0.64
FPS	FEAR 2: Project Origin	PS3	79	0.34

FPS	Call of Duty 3	PS3	80	0.7
FPS	The Chronicles of Riddick: Assault on Dark Athena	PS3	80	0.16
FPS	Tom Clancy's Rainbow Six: Vegas 2	PS3	81	1.2
FPS	Condemned 2: Bloodshot	PS3	82	0.31
FPS	Borderlands	PS3	83	0.82
FPS	Call of Duty: World at War	PS3	85	4.29
FPS	Far Cry 2	PS3	85	1.13
FPS	Resistance	PS3	86	3.71
FPS	Unreal Tournament III	PS3	86	0.57
FPS	Resistance 2	PS3	87	2.01
FPS	Battlefield: Bad Company 2	PS3	88	1.95
FPS	Bioshock 2	PS3	88	0.82
FPS	Killzone 2	PS3	91	2.42
FPS	Bioshock	PS3	94	0.72
FPS	Call of Duty: Modern Warfare 2	PS3	94	8.86
RPG	Operation Darkness	XBOX 360	46	0.03
RPG	Two Worlds	XBOX 360	50	0.48
RPG	Kingdom Under Fire: Circle of Doom	XBOX 360	55	0.32
RPG	Spectral Force 3	XBOX 360	59	0.07
RPG	Risen	XBOX 360	60	0.12
RPG	Divinity II: Ego Draconis	XBOX 360	62	0.14
RPG	Alpha Protocol	XBOX 360	63	0.19
RPG	Phantasy Star Universe	XBOX 360	64	0.1
RPG	Too Human	XBOX 360	65	0.72
RPG	Final Fantasy XI: Online	XBOX 360	66	0.22
RPG	The Last Remnant	XBOX 360	66	0.64
RPG	Nier	XBOX 360	67	0.14
RPG	Infinite Undiscovery	XBOX 360	68	0.6
RPG	Enchanted Arms	XBOX 360	69	0.19
RPG	Record of Agarest War	XBOX 360	71	0.14
RPG	Sacred 2: Fallen Angel	XBOX 360	71	0.43
RPG	Star Ocean: The Last Hope	XBOX 360	72	0.64
RPG	Marvel Ultimate Alliance 2	XBOX 360	73	0.74

RPG	Resonance of Fate	XBOX 360	74	0.2
RPG	Culdcept SAGA	XBOX 360	75	0.17
RPG	Lost Odyssey	XBOX 360	78	0.84
RPG	Blue Dragon	XBOX 360	79	0.56
RPG	Eternal Sonata	XBOX 360	79	0.25
RPG	Tales of Vesperia	XBOX 360	79	0.54
RPG	Final Fantasy XIII	XBOX 360	82	1.62
RPG	Marvel: Ultimate Alliance	XBOX 360	82	2.48
RPG	Dragon Age: Origins	XBOX 360	86	1.86
RPG	Fable II	XBOX 360	89	3.9
RPG	Mass Effect 2	XBOX 360	91	2.21
RPG	Fallout 3	XBOX 360	93	3.4
RPG	The Elder Scrolls IV: Oblivion	XBOX 360	94	3.43
RPG	Mass Effect	XBOX 360	96	2.32
RPG	Last Rebellion	PS3	44	0.05
RPG	Untold Legends: Dark Kingdom	PS3	58	0.13
RPG	Trinity Universe	PS3	62	0.09
RPG	Enchanted Arms	PS3	64	0.21
RPG	White Knight Chronicles: International Edition	PS3	64	0.68
RPG	Atelier Rorona: Alchemist of Arland	PS3	65	0.16
RPG	Record of Agarest War	PS3	67	0.03
RPG	Nier	PS3	68	0.31
RPG	Sacred 2: Fallen Angel	PS3	70	0.42
RPG	Alpha Protocol	PS3	72	0.19
RPG	Resonance of Fate	PS3	72	0.43
RPG	Marvel Ultimate Alliance 2	PS3	74	0.59
RPG	Star Ocean: The Last Hope	PS3	74	0.4
RPG	Folklore	PS3	75	0.21
RPG	3D Dot Game Heroes	PS3	77	0.26
RPG	Marvel: Ultimate Alliance	PS3	78	0.31
RPG	Eternal Sonata	PS3	80	0.17
RPG	Final Fantasy XIII	PS3	83	4.37
RPG	Valkyria Chronicles	PS3	86	0.94
RPG	Dragon Age: Origins	PS3	87	1

RPG	Demon's Souls	PS3	89	0.82
RPG	Fallout 3	PS3	90	2.37
RPG	The Elder Scrolls IV: Oblivion	PS3	94	1.99

References

- Baker, L. (2011) THQ shares fall on reviews of "Homefront" war game. Reuters, March 15, 2011. Available at: <http://www.reuters.com/article/2011/03/15/us-thq-shares-idUSTRE72E7E620110315> [Accessed: 9 May 2011].
- Brightman, J. (2009) Interview: John Riccitiello on E3, fighting piracy, metacritic, and more. Available at: <http://www.industrygamers.com/news/interview-john-riccitiello-on-e3-fighting-piracy-metacritic-and-more/3/> [Accessed: 8 February 2012].
- Boesky, K. (2008) Opinion: Why EA, the industry shouldn't rely on metacritic. Gamasutra, May 23, 2008. Available at: http://www.gamasutra.com/php-bin/news_index.php?story=18562 [Accessed: 15 August 2011].
- Dodson, J. (2006) Mind over meta. GameRevolution, July 14, 2006. Available at: http://www.gamerevolution.com/features/mind_over_meta [Accessed: 9 May 2011].
- Doyle, M. (2011) About Metacritic. Available at: <http://www.metacritic.org/about/> [Accessed: 27 September 2011].
- Dring, C. (2010) EA's Moore: Metacritic mania a slippery slope. Develop, 7/20/2010. Available at: <http://www.develop-online.net/news/35425/EAs-Moore-Metacritic-mania-a-slippery-slope> [Accessed: 18 August 2012].
- Everiss, B. (2008) Metacritic has changed the games industry. Available at: <http://www.bruceongames.com/2008/06/04/metacritic-has-changed-the-games-industry/> [Accessed: 8 February 2012].
- Fahey, M. (2011) Dragon Age II dev rates his own game on Metacritic, EA bets Obama voted for himself too. Kotaku, 3/15/2011. Available at: <http://kotaku.com/5782097/dragon-age-ii-dev-rates-his-own-game-on-metacritic-ea-bets-obama-voted-for-himself-too> [Accessed: 18 August 2012].
- Gilbert, B. (2012) Obsidian missed Fallout: New Vegas Metacritic bonus by one point. Available at: <http://www.joystiq.com/2012/03/15/obsidian-missed-fallout-new-vegas-metacritic-bonus-by-one-point/> [Accessed: 9 May 2011].
- Graft, K., Sheffield, B., Nutt, C., Rose, M., Cifaldi, F., Caoili, E., Alexander, L., Miller, P., Curtis, T. (2012) Ask Gamasutra: 84 Metacritic need not apply. Gamasutra, 7/24/2012. Available at: http://gamasutra.com/view/news/174829/Ask_Gamasutra_84_Metacritic_need_not_apply.php [Accessed: 20 August 2012].

- Greenwood-Ericksen, A. (2011) On the Role of Metacritic in the Game Industry. Available at: <http://www.fsoblogs.com/gdms/?currentPage=4> on 8/15/2011
- Lazarus, D. (2001) "San Francisco Chronicle, 2001". Sfgate.com, cited in Wikipedia. Available at: http://en.wikipedia.org/wiki/Rotten_Tomatoes 8/15/2011 [Accessed: 4 December 2009].
- McDonald, K. (2012) Is metacritic ruining the games industry? IGN, 6/16/2012. Available at: http://www.ign.com/articles/2012/07/16/is-metacritic-ruining-the-games-industry?utm_source=Monday%20newsletter&utm_medium=email&utm_campaign=7.17%20Dynamic%20Newsletter_NO%20FNAME_5949_280581_280628&utm_content=16600107 [Accessed: 18 August 2012].
- Metacritic. (2010a) I read Manohla Dargis' review of [MOVIE NAME] and I swear it sounded like a 90... why did you say she gave it an 80? Available at: https://metacritic.custhelp.com/app/answers/detail/a_id/1501/session/L3Nuby8wL3NpZC9DOFVxQkczaw== [Accessed 20 May 2012].
- Metacritic. (2010b) Can you tell me how each of the different critics are weighted in your formula? Available at: https://metacritic.custhelp.com/app/answers/detail/a_id/1507/session/L3Nuby8wL3NpZC9DOFVxQkczaw== [Accessed 27 May 2012].
- Metacritic. (2010c) Why don't you have 97 reviews for every movie like those other websites do? Available at: http://metacritic.custhelp.com/app/answers/detail/a_id/1510/session/L2F2LzEvdGltZS8xMzM5NDI2MTYyL3NpZC90anp0NnAtaw%3D%3D/sno/0 [Accessed 13 June 2012].
- Metacritic. (2012a) How we create the metascore magic. Retrieved 5/20/2012 from <http://www.metacritic.com/about-metascores> [Accessed: 20 May 2012].
- Metacritic. (2012b) Which critics and publications are included in your calculations? Available at: https://metacritic.custhelp.com/app/answers/detail/a_id/1508/session/L3Nuby8wL3NpZC9DOFVxQkczaw== [Accessed: 13 June 2012].
- Murdoch, J. (2010) Metacritic: Gaming the score. Gamepro. Available at: <http://www.gamepro.com/article/features/214841/gaming-the-score-metacritic/> [Accessed: 28 September 2011].
- Nutt, C. (2012) How Creative Assembly's process breeds quality. Gamasutra, 4/30/2012. Available at: http://gamasutra.com/view/feature/169354/how_creative_assemblys_process.php [Accessed: 30 April 2012].
- O'Rourke, K. (2007) "An historical perspective on meta-analysis: dealing quantitatively with varying study results". Journal of the Royal Society of Medicine, 100 (12): 579–582. doi:10.1258/jrsm.100.12.579. PMID 18065712.

- Periera, C. (2012) OP-ED: Metacritic presents real problems for the industry. 1UP, 7/16/2012. Available at: <http://www.1up.com/news/metacritic-presents-problems-industry> [Accessed: 18 August 2012].
- Pham, A., & Fritz, B. (2011) Bad reviews of Homefront send THQ shares tumbling. The Los Angeles Times, March 16, 2011. Available at: <http://articles.latimes.com/2011/mar/16/business/la-fi-ct-thq-homefront-20110316> [Accessed: 9 May 2012].
- Sinclair, B. (2011) Jurassic Park user reviews abused. Gamespot, 11/17/2012. Available at: <http://uk.gamespot.com/features/jurassic-park-user-reviews-abused-6346288/?tag=updates%3Beditor%3Ball%3Btitle%3B2> [Accessed: 18 August 2011].
- Wingfield, N. (2007) High scores matter to game makers, too. The Wall Street Journal, September 20, 2007. Available at: http://online.wsj.com/public/article/SB119024844874433247-EnpxM1F6fI9YZDofC7VnyPzVrGQ_20070920.html [Accessed: 15 August 2012].