

Measuring the Aesthetic Success of Books: Can User-driven Databases Fill the Gap?

Sven Form

Kiel University, Germany

E-mail address: form@psychologie.uni-kiel.de

ARTICLE INFO

Keywords:

Indicator of success
Books
Criterion problem
Bestseller
Aesthetic success
Creativity judgment

Article history:

Received 18 August 2017
Received in revised form 24 October 2017
Accepted 3 November 2017
ISSN: 2354-0036
DOI: 10.1515/ctra-2017-0016

ABSTRACT

Researchers that wish to evaluate the aesthetic success or functional creativity of books in the real world need a method to measure the outcome variable. However, sales figures are rarely published. Bestseller lists and expert judgments may not adequately reflect the aesthetic success among the general public. Data available on the platform Goodreads may serve as an alternative for measuring the popularity of books. In the present study, the ratings and number of ratings from Goodreads, as well as the number of literary prizes awarded are compared with the actual number of copies sold for a significant sample, the 98 most bestselling books in the UK from 1998 to 2012. Results indicated that literary prizes cannot serve as a gauge for the popularity of a book. While ratings were associated with copies sold, the number of ratings was a significantly better indicator of the sales figures of a book.

INTRODUCTION

Whether a book becomes aesthetically successful is influenced among other factors by the creativity of the work. But to actually *determine* the exact contribution of a product's creativity, one needs also to measure the outcome variable aesthetic success.

Generalizing considerations from the domain of music (Simonton, 1998), two kinds of criteria can be distinguished for the aesthetic success of creative products: subjective and objective ratings. Subjective ratings reflect judgments by some individuals or even a single individual in a laboratory setting. Objective ratings measure the frequency with which the creative product is heard, watched, staged, - or in the present domain of interest - read in the real world. Simonton (1998) considered this objective rating to reflect popularity.

However, when researchers try to determine which properties of a literary work influence how a book is judged and ultimately becomes aesthetically successful, the resulting dependent variable is not easy to measure outside the laboratory. Using economic data as an indicator for the recognition of a book is not as straightforward as it seems. The actual number of copies sold is rarely known, as publishing houses shy away from making their numbers official. Accordingly, it is not surprising that there are only a handful of studies, which have been based on the number of sold copies (Beck, 2006; Clerides, 2002; Form, 2017; Schmidt-Stölting, Blömeke, & Clement, 2011; Sorensen, 2007).

Using published bestseller lists, as an alternative indicator, has the disadvantage that they offer only relative rankings. This does not reflect the substantial disparity in the success of books, that is, success of books is right-skewed (Schmidt-Stölting et al., 2011), like the fame of poets (Martindale, 1995). Furthermore, the criteria for inclusion on such lists are sometimes surprisingly arbitrary, leading, for example, to the exclusion of the *Harry Potter* series from the *Publishers Weekly* list. A final argument against using bestseller lists is they consider only a certain period of time, making it difficult to compare books published over a long time range.

With internet platforms like Goodreads, a researcher may ask whether the freely available information on such user-driven databases can serve as an indicator of books' success in the absence of other "hard figures". Similarly to IMDB for movies, users can search a database of books on Goodreads (website at <http://www.goodreads.com/>). They can also give ratings for any book that is registered in the database which listed 395 million in 2012 (Fidelman, 2012). Using the data from Goodreads could be especially promising, as Goodreads had 20 million registered members in 2013 (Chandler, 2013), suggesting a large number of laypersons' judgments lying unused under the surface of the platform.

However, one could argue that a layperson's judgment is not necessary, because the judgment made by experts (measured outside the laboratory, for example, by prizes awarded [Kaufman, Baer, Cole, & Sexton, 2008]) should be sufficient to evaluate creativity. In fact, the judgment by experts, called the Consensual Assessment Technique (Amabile, 1982), is considered the gold standard (Kaufman, Plucker, & Baer, 2008) and strongly recommended (Kaufman & Baer, 2012) for assessing creativity.

The distinction between nonexpert and expert judgment seems to be especially relevant to the domain of writing, where critics regularly shake their heads when they see which books make it onto the bestseller lists. Indeed, empirical results from laboratory studies in the domain of writing have indicated, that the level of expertise influences

the degree of agreement between judges. Published poets agreed more in their judgment than college students when judging poems (Kaufman, Baer, et al., 2008), as in the case of writers compared to students judging short stories (Kaufman, Baer, & Cole, 2009). However, writers as experts did not agree more than “quasi-experts”, namely research students of creativity, teaching students, and English teachers when judging those short stories (Kaufman, Baer, Cropley, Reiter-Palmon, & Sinnott, 2013). Similarly, librarians, who can also be considered as quasi-experts, had a level of agreement comparable to English lecturers when judging book excerpts (Nell, 1988b). Experts, or at least quasi-experts, seem to be a good choice for judging creative products in the domain of writing (Kaufman & Baer, 2012).

Nonetheless, creativity researchers do not always exclude the nonexpert perspective when assessing creativity outside the laboratory. In fact, being recognized not only by some specialized experts, but by the public (e.g. in nationwide publications) is considered a higher contribution to a field in the stepwise-ranked items of the Creative Achievement Questionnaire, which is a widely used measure for an individual's creativity (Carson, Peterson, & Higgins, 2005; Form, Schlichting, & Kaernbach, 2017).

But rather than considering one kind of recognition as a higher achievement than another, it is probably more fruitful to understand the recognition of books by taking a look at the domain of cinematic creativity. Here, creative products are often used to entertain and to make profits, similarly to the domain of writing. Recognition in this domain has been loosely divided into two aspects (Simonton, 2005, 2009b): popularity, measured in terms of economic success, and experts' praise, e.g. assessed by critical acclaim. For example, Simonton (2007) used both critical acclaim and economic success as dependent variables to measure the recognition of cinematic creativity. This approach was justified, as this differentiation made it possible to determine that different factors lead to either popularity or critical acclaim (Simonton, 2005). Although critical acclaim had a positive effect on the economic success of movies (Plucker, Holden, & Neustadter, 2008; Simonton, 2009a) [which is another parallel to the domain of books (Clement, Proppe, & Sambeth, 2006; Keuschnigg, 2012; Schmidt-Stölting et al., 2011)], success was even negatively correlated with critical acclaim (Simonton, 2005).

The fact that public perception must be taken into account, at least in some domains, can be deduced from the very core of how creativity is defined. It rests on the two pillars of originality (or novelty) and effectiveness (or usefulness) (Runco & Jaeger, 2012). Often, creativity researchers implicitly focus their interest on the aspect of originality (Gilson & Madjar, 2011), probably because it is a first-order criterion of creativity (Diedrich, Bene-

dek, Jauk, & Neubauer, 2015). But what about cases in which the effectiveness of the creative product merely lies in entertaining a lot of people or providing a nice read before falling asleep?

As Kaufman et al. (2012) pointed out, it is important to consider for whom a creative product is new and useful. In the real world, effectiveness and usefulness are usually preferred over originality (Cropley & Cropley, 2008). For laypersons, high originality is generally not only of lesser relevance, they even prefer less original products (Blair & Mumford, 2007; Rietzschel, Nijstad, & Stroebe, 2010; Staw, 1995). This is probably the reason why, even those genre novels, which can be categorized as *replications* (Sternberg, Kaufman, & Pretz, 2001) and thus lack high levels of originality, can be very successful. Obviously, they are sufficiently “effective” in their purpose. If a researcher wants to investigate works of such *functional creativity* (Cropley & Cropley, 2005), they will need indicators of popularity. With respect to the domain of cinematic creativity, it is however known, that expert judgments poorly represent popularity (Simonton, 2005).

As popularity is the target variable, the present study investigated, first, whether expert judgments (as measured in literary prizes awarded) are a valid indicator of popularity (as measured in terms of economic success). Second, it was determined whether ratings, and the number of ratings from the internet-based platform Goodreads, are associated closely enough with economic success to serve as an alternative measure of popularity, in other samples of books for which economic data are not available. From the evaluation of user-rating data in the field of cinematic creativity, it is however known, that the number of reviews serves as a slightly better predictor of economic success than the reviews themselves (Plucker et al., 2008). Thus, it was also determined whether ratings themselves, and the number of ratings accrued, differ in their association to economic success.

METHOD

The numbers for copies sold was taken from a freely available online article on the website of *The Guardian* (“The top 100 bestselling books of all time,” 2012). The original list covered 100 books, which sold the most copies in the UK from 1998 to 2012 ($M = 1,519,496$ copies; $Mdn = 1,111,939$ copies) (see Table 1), thus representing what Simonton (2014) called a significant sample. This list also includes ten books, which were first published in the UK before 1998. Overall, the year of their first publication in the UK ranged from 1955-2012 ($M = 2002.6$, $Mdn = 2004$). The survey covered about 6,500 retailers and over 90% of all retail book purchases in the UK (“Top-selling 100 books of all time,” 2011).

The present sample included 98 instead of 100, because one book represented the “children's edition” of another book on the list, *Harry Potter and the Half-Blood Prince*.

The two versions of the books differed only in their cover. Accordingly, sales of both editions were added together. The other case removed from the list was *The Annual Beano*, as it represents not one book, but the summed up sales of a series of books which have been published every year. The sample of 98 books was written by 66 authors, 51.5% of British, 30.3% of US-American, 4.6% of Irish and 13.6% of other origin. Forty-four percent of the authors were females.

Ratings and number of ratings were taken from the web page of the respective book on Goodreads on the 23th and 24th May, 2017. Ratings on Goodreads can theoretically range from 1 to 5. As the sales numbers and the number of ratings showed right-skewness, both were log-transformed.

To have a gauge for expert ratings, several book awards were used to sum up an expert score. For each source the following scheme was used: 1 = recipient, 0 = no award. The expert score included from the *British Book Awards* the categories: *Children's Book of the Year*, *Biography of the Year*, *Popular Fiction Book of the Year*, *Popular Non-Fiction Book of the Year*, *Thriller & Crime Novel of the Year*, and *Food & Drink Book of the Year*. From the additional category *Book of the year*, it included only laureates up to 2009, as the public has been choosing the laureate since 2010, which does not reflect expert ratings. Furthermore, the expert score included the *Man Booker Prize*, the *Costa Book Award*, the *Baillie Gifford Prize*, the *Guardian First Book Award*, the *James Tait Black Prize* and the *Betty Trask Award*.

All analyses were performed with SPSS Statistics 20.0 software.

Table 1
Descriptive Statistics of Untransformed Variables and Correlations
Between Transformed Variables

	Min	Max	<i>M (SD)</i>	<i>Mdn</i>	1	2	3	4
Sold copies ¹ (1)	791,095	5,094,805	1,519,496 (939,543)	1,111,939				
Rating (2)	3.30	4.61	3.94 (.31)	3.93	.28**			
Number of ratings ¹ (3)	333	4,636,854	598,798 (860,050)	207,222	.58**	.30**		
Year of 1 st publishing (4)	1955	2012	2002.6 (8.4)	2004	.10	-.22*	-.11	
Number of awards ¹ (5)	0	3	.3 (.54)	0	.16	-.11	.10	.10

Note: N = 98. 1 variable was log-transformed for correlation analysis. *p < .05, **p < .005

RESULTS

To test which of the alternative measures for popularity was most similar to economic success, correlational analyses were performed (see Table 2). The number of literary awards a book won was not associated with its economic success. Ratings ($r = .28$, $p = .005$) and number of ratings ($r = .58$, $p < .001$) were both significantly correlated with the number of sold copies (Table 2). A principal component analysis on the three variables with oblique rotation indicated one single factor with an eigenvalue above 1, explaining 59.8% of variance, suggesting popularity as an underlying latent variable.

The number of ratings and ratings themselves correlated with the number of copies sold in different magnitudes ($r = .58$ vs. $r = .28$). To test whether this difference was significant, a comparison of correlations was performed using Hittner et al.'s (2003) evaluation method as implemented in the web interface by Diedenhofen and Musch (2015). The comparison of correlations indicated that the number of ratings was significantly better associated with economic success than the ratings themselves ($z = -2.89$, $p = 0.002$).

One could argue that the association between the number of ratings and the number of copies sold is merely attributable to confounding variables like a preference for authors of a certain origin or the year of first publication. For example, both ratings given and copies sold accumulate over time which could lead to a spurious association between them. Thus, a regression analysis was performed using number of copies sold as independent variable and year of first publication and origin of author as control variables (dummy-coded with US-American origin as reference) (Table 3). Only the British origin of an author had an effect on the number of ratings as the dependent variable, namely a negative one.

Table 2
Regressions for Number of Ratings on Number of Copies Sold

Variable	log(Number of ratings)							
	Model 1				Model 2			
	B	SE	β	p	B	SE	β	p
Log(Number of copies sold)	2.8	.41	.58	<.001	3.0	.37	.61	<.001
Covariates								
Year 1 st published					-.02	.01	-.17	.02
British author					-.66	.17	-.33	<.001
Other origin					.25	.24	.09	.29
Adjusted R^2	.33				.48			

Note: N = 98. "Other origin" refers to authors of neither British, nor US-American authorship.

DISCUSSION

The aim of the present study was to compare different measures of the success of books using sales numbers as an alternative for measuring the popularity of books. Specifically, data from the database Goodreads were considered.

As hypothesized, economic success is not well reflected by the number of awards a book has won. Obviously, while literature awards have a beneficial effect at least on sales of hard covers (Schmidt-Stölting et al., 2011), this relationship is not sufficiently strong to reflect the taste of the general audience. These findings parallel those from cinematic creativity (Plucker et al., 2008; Simonton, 2005, 2009a), but contrast with results from highbrow art (Pénet & Lee, 2014).

The comparison of correlations indicated that the number of ratings on Goodreads is a significantly better predictor of economic success than the ratings themselves. Although being written by a British author had a negative effect on the number of ratings compared to US-American authorship, the association between the number of ratings and the number of copies sold remained. As British sales numbers were able to predict the popularity among (probably) mostly US-American users, the association to sales on the US book market should accordingly be even closer.

An aspect of the study should be considered that might be seen as limiting its scope and conclusions. A challenge was that sales numbers in the present sample were from the UK book market, while Goodreads seems to be rather oriented to US-American users. (The latter is suggested because books titles used in the database are the titles under which books were published in the US, rather than the British titles. Notwithstanding the fact that British users can also rate books on the website.) Thus, a question here was: Can the economic success of books in the UK predict rating data on a US-American oriented website? At first sight, it would appear that such a cross-national comparison might hamper the meaningful use of data. Nevertheless, there are several reasons why this approach could work. First, as in the present sample, only top bestselling books were considered, and their availability should be comparable in both countries. Second, individuals generally agree as regards the judgment of art and other stimuli across cultures (Berlyne, 1975, 1976; Child, 1981). Specifically for books, reading for pleasure is probably relatively invariant across taste and even national cultures, as the “appetite for narrative” is universal (Nell, 1988a, 1988b). This view is supported by a preference for the same written narrative even when the narrative’s origin is from a foreign culture (Morra & Lazzarini, 2002). Third, not only was the content of the books unchanged in the present sample, but also the form, since they did not have to be translated. Fourth, despite all likely noise,

there was a latent variable detectable affecting both data sets. This makes a comparison feasible. Fifth, we could control for possibly confounding variables (like year of publication) or biasing variables (such as origin of author).

If one wants to measure the popularity of a book and compares the validity of the data from Goodreads with the number of copies sold, one could even argue that the data from Goodreads are in fact more valid: many copies are bought, but are merely given away as presents without being read or end up unread on book shelves. Other copies will be read by more than one person, if a book is lent to others. Both cases can be considered as measurement error when relying on the measure of number of copies sold. For the data from Goodreads this kind of “noise” probably does not exist. To be rated, a book has to be read.

In summary, when numbers about copies sold are not available, the use of literary prizes awarded as an alternative indicator of popularity is discouraged, and to a lesser extent, so is using ratings of books on Goodreads. However, the number of ratings appear to be a valid substitute for the measure of number of copies sold, especially when the origin of the author is controlled for. Future studies can now use the data to determine which properties of a book influence its success as a creative product.

REFERENCES

- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997-1013.
- Beck, J. (2006). The sales effect of word of mouth: A model for creative goods and estimates for novels. *Journal of Cultural Economics*, 31, 5-23.
- Berlyne, D. E. (1975). Extension to Indian subjects of a study of exploratory and verbal responses to visual patterns. *Journal of Cross-Cultural Psychology*, 6, 316-330.
- Berlyne, D. E. (1976). Similarity and preference judgments of Indian and Canadian subjects exposed to Western paintings. *International Journal of Psychology*, 11, 43-55.
- Blair, C. S., & Mumford, M. D. (2007). Errors in idea evaluation: Preference for the unoriginal? *Journal of Creative Behavior*, 41, 197-222.
- Carson, S., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17, 37-50.
- Chandler, O. (2013, July 23). Goodreads grows to 20 million readers [Web log message]. Retrieved from <https://www.goodreads.com/blog/show/425-goodreads-grows-to-20-million-readers>
- Child, I. L. (1981). Bases of transcultural agreement in response to art. In *Advances in Intrinsic Motivation and Aesthetics* (pp. 415–432). Boston, MA: Springer US.

- Clement, M., Proppe, D., & Sambeth, F. (2006). Der Einfluss von Meinungsführern auf den Erfolg von hedonischen Produkten [The impact of opinion leaders on the success of hedonic products]. *Journal of Business Economics*, 76, 797-824.
- Clerides, S. K. (2002). Book value: Intertemporal pricing and quality discrimination in the US market for books. *International Journal of Industrial Organization*, 20, 1385-1408.
- Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10(4), 1-12.
- Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, 9, 35-40.
- Form, S. (2017). Reaching wuthering heights with brave new words: The influence of originality of words on the success of paramount bestsellers. *Journal of creative behavior*, accepted.
- Form, S., Schlichting, K., & Kaernbach, C. (2017). Mentoring functions: Interpersonal tensions are associated with mentees' creativity. *Psychology of Aesthetics, Creativity, and the Arts*, accepted.
- Fidelman, M. (2012, October 16). These are top 25 book reviewers on Goodreads (infographic). *Forbes*. Retrieved from <http://www.forbes.com/sites/markfidelman/2012/10/16/goodreads-ceo-these-top-25-book-reviewers-represent-the-future-infographic/?ss=strategies-solutions>
- Gilson, L. L., & Madjar, N. (2011). Radical and incremental creativity: Antecedents and processes. *Psychology of Aesthetics, Creativity, and the Arts*, 5, 21-28.
- Hittner, J., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, 130, 149-168.
- Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, 24, 83-91.
- Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the Consensual Assessment Technique. *The Journal of Creative Behavior*, 43, 223-233.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the Consensual Assessment Technique. *Creativity Research Journal*, 20, 171-178.
- Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., & Sinnott, S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts*, 7, 332-340.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. (N. L. K. Alan S. Kaufman, Ed.), *Essentials Psychological Assessment Series*. Hoboken, NJ: John Wiley & Sons, Inc.

- Keuschnigg, M. (2012). *Das Bestseller-Phänomen: Die Entstehung von Nachfragekonzentration im Buchmarkt [The bestseller phenomenon: The emergence of demand concentration in the book market]*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Martindale, C. (1995). Fame more fickle than fortune: On the distribution of literary eminence. *Poetics*, 23, 219-234.
- Morra, S., & Lazzarini, S. (2002). A cross-cultural study of response to Icelandic and Italian folktales. *Empirical Studies of the Arts*, 20, 61-82.
- Nell, V. (1988a). *Lost in a book: The psychology of reading for pleasure*. New Haven: Yale University Press.
- Nell, V. (1988b). The psychology of reading for pleasure: Needs and gratifications. *Reading Research Quarterly*, 23, 6-50.
- Pénet, P., & Lee, K. (2014). Prize & price: The Turner Prize as a valuation device in the contemporary art market. *Poetics*, 43(1), 149-171. <https://doi.org/10.1016/j.poetic.2014.01.003>
- Plucker, J. a., Holden, J., & Neustadter, D. (2008). The criterion problem and creativity in film: Psychometric characteristics of various measures. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 190-196.
- Rietzschel, E., Nijstad, B. A., & Stroebe, W. (2010). The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology*, 101, 47-68.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24, 92-96.
- Schmidt-Stölting, C., Blömeke, E., & Clement, M. (2011). Success drivers of fiction books: An empirical analysis of hardcover and paperback editions in Germany. *Journal of Media Economics*, 24, 24-47.
- Simonton, D. K. (1998). Masterpieces in music and literature: Historiometric inquires. *Creativity Research Journal*, 11(2), 103-110. https://doi.org/10.1207/s15326934crj1102_2
- Simonton, D. K. (2005). Film as art versus film as business: Differential correlates of screenplay characteristics. *Empirical Studies of the Arts*, 23, 93-117.
- Simonton, D. K. (2007). Film music: Are award-winning scores and songs heard in successful motion pictures? *Psychology of Aesthetics, Creativity, and the Arts*, 1, 53-60.
- Simonton, D. K. (2009a). Cinematic success, aesthetics, and economics: An exploratory recursive model. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 128-138.
- Simonton, D. K. (2009b). Cinematic success criteria and their predictors: The art and business of the film industry. *Psychology & Marketing*, 26, 400-420.

- Simonton, D. K. (2014). Significant samples-not significance tests! The often overlooked solution to the replication problem. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 11-12.
- Sorensen, A. T. (2007). Bestseller lists and product variety. *The Journal of Industrial Economics*, 55, 715-738.
- Staw, B. M. (1995). Why no one really wants creativity. In C. M. Ford & D. Gioia (Eds.), *Creative action in organizations* (pp. 161-166). Thousand Oaks, CA: Sage Publications.
- Sternberg, R. J., Kaufman, J. C., & Pretz, J. E. (2001). The propulsion model of creative contributions applied to the arts and letters. *The Journal of Creative Behavior*, 35, 75-101.
- The top 100 bestselling books of all time: How does Fifty Shades of Grey compare? (2012, August). *The Guardian*. Retrieved from <https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-time-fifty-shades-grey-compare>
- Top-selling 100 books of all time. (2011, January). *The Guardian*. Retrieved from <https://www.theguardian.com/news/datablog/2011/jan/01/top-100-books-of-all-time#data>

Corresponding author at: Sven Form, Department of Psychology, Olshausenstraße 62, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany.
E-mail: form@psychologie.uni-kiel.de

