- Accessibility: concerns whether the user has access to an information system in order to meet her information needs. Its dimensions include accessibility and access security.

Pipino and colleagues [16] noted that the framework can be used in an objective assessment of quality in particular contexts. Metrics should be developed that operationalize the quality dimensions relevant to the data set and task at hand. For example, a previous study used the framework to predict quality in news articles [23]. Textual properties such as length and the presence of key vocabulary were found to correlate to aspects of quality.

We determined that to assess quality in Amazon reviews, only the first three categories in the framework are needed. Accessibility is not relevant since participants in the community are using the same information system (i.e. the virtual community environment). Table 3 shows the quality framework developed for the current study. As can be seen, we have incorporated 9 aspects of quality across the first three categories. The third column of Table 3 describes the metrics used to operationalize the dimensions of quality. We have incorporated information from four sources: the textual properties of the reviews (e.g. length, vocabulary), metadata of the reviews (e.g. age), information from the respective reviewer's Amazon profile, and properties of the products themselves

Category	Dimensions	Metrics	Explanation / Justification
Intrinsic quality	Accuracy Objectivity	Textual similarity between the review and description on product's page. In particular, the (1) cosine, (2) bigram overlap, and (3) normalized longest common subsequence between the two texts were calculated [14].	[6] proposed that there are two types of information in reviews: objective, which is textually similar to the product description, and subjective, that differs from the description.
	Believability Reputation	(4) Product rating (on a 5-point scale) assigned by reviewer (5) Reviewer uses real name (6) Reviewer has top reviewer badge (7) Reviewer's rank in the community (8) Total reviews contributed by reviewer (9) # Helpful votes received by reviewer (10) Perplexity of textual review (11) Entropy of textual review	(4): Consumers with extreme opinions of a product are more likely to write reviews and often want to vent their frustrations [1]. (5)-(9): These attributes might be used by community members to assess reviewer reputation. (10)-(11): If we consider the distribution of words used in all reviews of a product, perplexity and entropy quantify the deviation of a review from what is expected [14].
Contextual quality	Relevancy	(12) Centroid (textual centrality) score of product review, as described in [17].	A weighted vector of words used across all reviews of a product is created. A review's centroid score quantifies the extent to which it contains words that are statistically important across reviews.
	Appropriate amount	Length of review measured as: (13) # Sentences (14) # Words	Trivially, longer texts contain more information. However, some reviews could be too long for users to read.
	Timeliness	(15) Days lapsed since the earliest review was posted about the respective product	Older reviews tend to have fewer ratings [4, 15].
Representational quality	Ease of understanding Interpretability	"Readability" measures of review: (16) Characters-to-sentences ratio (17) Words-to-sentences ratio	Texts that score high on these measures are more complex and take more effort to understand [3].

Table 3: Wang and Strong's (1996) data quality categories, dimensions and the metrics used to quantify them.

TABLE 1
The Variables Collected for Our Study

Type	Variable	Explanation
Product and Sales Data	Retail Price	The retail price at Amazon.com
	Sales Rank	The sales rank within the product category
	Average Rating	Average rating of the posted reviews
	Number of Reviews	Number of reviews posted for the product
	Elapsed Date	Number of days since the release of the product
Individual Review	Moderate Review	Does the Review have a moderate rating (3 star rating) or not
	Helpful Votes	The number of helpful votes for the review
	Total Votes	The total number of votes for the review
	Helpfulness	$\frac{Helpful Votes}{Total Votes}$
Reviewer Characteristics	Reviewer Rank	The reviewer rank according to Amazon
	Top-10 Reviewer	Is the reviewer a Top-10 reviewer?
	Top-50 Reviewer	Is the reviewer a Top-50 reviewer?
	Top-100 Reviewer	Is the reviewer a Top-100 reviewer?
	Top-500 Reviewer	Is the reviewer a Top-500 reviewer?
	Real Name	Has the reviewer disclosed his/her real name?
	Nickname	Does the reviewer have a nickname listed in the profile?
	Hobbies	Does the reviewer have an "about me" section in the profile?
	Birthday	Does the reviewer list his/her birthday?
	Location	Does the reviewer disclose its location?
	Web Page	Does the reviewer have a home page listed?
	Interests	Does the reviewer list his/her interests?
	Snippet	Does the reviewer has a description in the reviewer profile?
	Any Disclosure	Does the reviewer list *any of the above* in the reviewer profile?
Reviewer History	Number of Past Reviews	Number of reviews posted by the reviewer
	Reviewer History Macro	Average past review helpfulness (macro-averaged)
	Reviewer History Micro	Average past review helpfulness (micro-averaged)
	Past Helpful Votes	Number of helpful votes accumulated in the past from the reviewer
	Past Total Votes	Number of total votes on the reviews posted in the past for the reviewer
Review Readability	Length (Chars)	The length of the review in characters
	Length (Words)	The length of the review in words
	Length (Sentences)	The length of the review in sentences
	Spelling Errors	The number of spelling errors in the review
	ARI	The Automated Readability Index (ARI) for the review
	Gunning Index	The Gunning–Fog index for the review
	Coleman–Liau index	The Coleman–Liau index for the review
	Flesch Reading Ease	The Flesch Reading Ease score for the review
	Flesch–Kincaid Grade Level	The Flesch–Kincaid Grade Level for the review
	SMOG	The Simple Measure of Gobbledygook score for the review
Review Subjectivity	AvgProb	The average probability of a sentence in the review being subjective
	DevProb	The standard deviation of the subjectivity probability

The panel data set contains data collected over a period of 15 months; we collected the variables daily and we capture the variability over time for the variables that change over time (e.g., sales rank, price, reviewer characteristics, and so on).

increase in sales for products, although the estimate is statistically significant only for audio-video players and digital cameras (see Table 5). It is statistically insignificant for DVDs. Our conjecture is that customers prefer to read reviews that describe the individual experiences of other consumers and buy products with significant such (subjective) information available only for *search goods* (such as cameras and audio-video players) but not for *experience goods.*[8]

The coefficient of *DevProb* has a positive and statistically significant relationship with sales rank in audio-video players and DVDs, but is statistically insignificant for digital cameras. In general, this suggests that a decrease in the deviation of the probability of subjective comments leads to a decrease in sales rank, i.e., an increase in product sales. This means that reviews that have a mixture of objective, and highly subjective sentences have a negative effect on product sales, compared to reviews that tend to include only subjective or only objective information.

8. Search goods are those whose quality can be observed before buying the product (e.g., electronics) while for experience goods, the consumers have to consume/experience the product in order to determine its quality (e.g., books, movies).

The coefficient of the *Readability* is negative and statistically significant for digital cameras suggesting that reviews that have higher Readability scores are associated with higher sales. This is likely to happen if such reviews are written in more authoritative and sophisticated language which enhances the credibility and informativeness of such reviews. Our results are robust to the use of other Readability

TABLE 2
Descriptive Statistics of Audio and Video Players for Econometric Analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
Retail Price	5699	151.33	130.57	0	3299.99
Sales Rank	7352	7667.42	51039.42	0	2090308
Log (Elapsed Date)	7352	5.12	1.09	0	7.63
Average Rating	7352	3.86	1.41	1	5
Number of Reviews	7352	195.07	138.76	0	522
Moderate Review	7352	0.093	0.29	0	1
Any Disclosure	7352	0.52	0.49	0	1
Helpful Votes	7352	5.51	11.7	0	744
Total Votes	7352	8.38	14.05	0	893
Log(Spelling Errors)	7352	-3.85	0.74	-6.67	-1.34
Readability (Gunning)	7352	12.46	13.31	1.36	277.95
AvgProb	7352	0.58	0.05	0.37	0.83
DevProb	7352	0.047	0.024	0	0.18
Rev. History Macro	3076	0.69	0.23	0	1