4. Beyond Relevance: Improving Documentation Quality with the Kano Model

Yoel Strimling
Ceva, Inc. Ra'anana, Israel

Abstract: This pilot study refines my previously proposed (Strimling, 2019) reader-derived definition of documentation quality (DQ) by applying the Kano Model of customer satisfaction to it. The aim is to develop a more reliable way to collect DQ feedback and metrics, as well as provide evidence-based resources for teaching students how to write quality documentation. Technical documentation readers were asked to rate the same DQ dimensions used in my previous study (and originally based on Wang and Strong's (1996) comprehensive information quality category/dimension framework) using a Kano Model questionnaire. The results indicate that the Accurate, Easy to Understand, and Accessible DQ dimensions continue to strongly align with their respective information quality categories: Intrinsic, Representational, and Accessibility. However, unlike in my previous study, the Complete DQ dimension emerged as a better representative of the Contextual information quality category than the Relevant DQ dimension. These findings strengthen my proposed reader-oriented DQ framework, with the Kano Model providing insights into how documentation features influence user satisfaction. This approach offers technical communication professionals and educators a foundation for developing robust feedback mechanisms, metrics, and teaching resources, and represents an initial attempt to apply the Kano Model's well-established methodology to DQ.

Keywords: Documentation quality, documentation quality metrics, documentation quality feedback, teaching documentation quality, Kano Model of customer satisfaction

Collecting meaningful and actionable feedback from readers about the technical documentation they use is an important tool that technical communicators can use to improve the quality of their content. This feedback can also be used to create consistent and reliable documentation quality (DQ) metrics that technical communicators can present to their managers to both prove the value of good documentation as well as determine where more effort needs to be invested. But collecting feedback about DQ and creating metrics that measure it are not easy tasks:

- It can be very difficult to get any kind of feedback at all from readers about

DOI: https://doi.org/10.37514/TPC-B.2026.2869.2.04

the documentation they use. There is often a disconnect between information producers and information consumers; this might be due to a lack of time, resources, or interest—or even a company policy of not asking customers about DQ (possibly because they are afraid of what readers will say and be unable to fix the problems they raise).

- Knowing what metrics we should be looking at to determine DQ can be very complicated. What should we be measuring? What is a useful metric and what is not? There are any number of DQ and usability criteria that we can collect—how do we know what is important? And how many metrics do we need to get a complete picture?

How to measure DQ is a major concern for technical communicators and their managers. In 2018, the Center for Information-Development Management (Stevens et al.) did a study to determine which critical metrics (for example, DQ, customer satisfaction, on-time delivery, and translation costs) their members were tracking and what they did with the information they collected. Most of them reported that they only tracked metrics that were easy to collect (what they called "metrics of convenience rather than metrics of significance"), and very few of them had any formal user feedback mechanism for collecting information from readers about the quality of their documentation.

According to their study, 76% of documentation managers were held accountable for DQ, but only 44% were measuring it in some way, and there was no guarantee that the metrics that they were collecting reflected how readers felt about DQ. Quite the contrary, in fact—while the quality metrics that were most measured were technical accuracy, clarity, and completeness, most respondents (86%) said that they relied mainly on editorial or peer reviews (not user reviews; this was only 34%) as the most common method of determining DQ. This begs the question—how can you possibly create meaningful metrics for these critical aspects of DQ without asking the actual documentation users?

Dawn Stevens et al. (2018) also found that 60% of their respondents were starting to turn to Web analytics (for example, views per page or time spent on a page) to collect additional insights about the quality of their documentation. User experience (UX) expert Jared Spool has spoken extensively about possible issues with using these types of analytics to measure quality. For example, Spool (2015; 2018) says that analytics like time on page or bounce rate, while easy to collect and track, do not really tell us much at all because there can be any number of different reasons for their values, and we can often infer contradictory deductions from them. It is possible to measure any number of criteria and track them over time to see increases or decreases—but without knowing if they are useful metrics or not, it is a waste of time to measure them. The only way to know if a metric is useful is to determine if it improves our users' experience—in other words, it is only our users who can tell us if we are measuring and tracking the right things. For DQ, this means that we must listen to the "voice of our readers" to find out

what they want from the documentation we send them. But to do this, we must first understand what our readers mean when they talk about DQ so we can align ourselves with their expectations. Only after we know how readers define DQ can we measure and track that to ensure that we are meeting their needs.

In this chapter, I will evaluate the preliminary framework for a focused, clearly defined, and reader-derived definition of DQ that I proposed in previous research (Strimling, 2019). That research was based on empirically tested and distinct information quality categories and dimensions presented in Richard Wang and Diane Strong (1996). In my 2019 study, I found that readers define high-quality documentation as being "accurate, relevant, easy to understand, and accessible," and proposed that this definition of DQ could be used to both collect meaningful and actionable feedback as well as provide clear and reliable metrics for measuring DQ.

I will attempt to make this definition of DQ from the readers' point of view more robust and strengthen its theoretical underpinnings by presenting the results of a pilot study I ran with the Kano Model of customer satisfaction. With a stronger definition of DQ, I think that a more reliable and comprehensive approach to DQ feedback and metrics can be formed. I have attempted to follow the recommendations presented by Rebekka Andersen and JoAnn Hackos (2018) to ensure that academic research findings are clearly applicable and relevant to technical communication practitioners. The need to understand how readers define DQ, and how we can use this definition to measure and improve reader satisfaction, are real-world issues that have clear practical implications.

Evaluating the Proposed DQ Definition

In 1996, Wang and Strong proposed a "comprehensive, hierarchical framework of data quality attributes" that were important to data consumers. Their underlying assumption was that, to improve data quality, they needed to understand what data quality meant to data consumers; data quality cannot be approached intuitively or theoretically because these do not truly capture the voice of the data consumer.

Their framework was made up of 15 quality dimensions, grouped into four quality categories: Intrinsic, Contextual, Representational, and Accessibility (ICRA), as listed in Table 4.1

Wang and Strong (1996) claimed that their proposed data quality framework of categories and dimensions could be used as a basis for further studies that measure perceived data quality in specific work contexts. They stated that the framework was methodologically sound, complete from the data consumers' perspective, and was useful for measuring, analyzing, and improving data quality. They cited "strong and convincing" anecdotal evidence that the framework had been used effectively in both industry and government and had helped data managers better understand their customers' needs by "identifying potential data deficiencies, operationalizing the measurement of these data deficiencies, and improving data quality along these measures" (p. 9).

Table 4.1. Data Quality Categories and Dimensions

Category	Dimensions
Intrinsic Quality: Data must have quality in its own right.	**Accuracy**: The data is correct, reliable, and certified free of error. **Believability**: The data is true, real, and credible. **Objectivity**: The data is unbiased (unprejudiced) and impartial. **Reputation**: The data is trusted or highly regarded in terms of its source or content.
Contextual Quality: Data must be considered within the context of the task at hand.	**The Appropriate Amount**: The quantity or volume of the available data is appropriate. **Completeness**: The data is of sufficient breadth, depth, and scope for the task at hand. **Relevance**: The data is applicable and helpful for the task at hand. **Timeliness**: The age of the data is appropriate for the task at hand. **Value**: The data is beneficial and provides advantages from its use.
Representational Quality: Data must be well represented.	**Conciseness**: The data is compactly represented without being overwhelming (that is, it is brief in presentation, yet complete and to the point). **Consistency**: The data is always presented in the same format and is compatible with previous data. **Ease of Understanding**: The data is clear, without ambiguity, and easily comprehended. **Interpretability**: The data is in an appropriate language and units, and the definitions are clear.
Accessibility Quality: Data must be easy to retrieve.	**Accessibility**: The data is available or easily and quickly retrievable. **Security**: Access to the data can be restricted, and hence, kept secure.

Source: Wang and Strong (1996)

While the originally stated object in Wang and Strong (1996) was *data quality*, it is now more commonly referred to as *information quality* (Arazy et al., 2017; Huang et al., 1999; Kahn et al., 2002; Lee et al., 2002; Pipino et al., 2002; Strong et al., 1997a, 1997b; Wang, 1998; Watts et al., 2009).

Wang and Strong's framework has been used as the basis for a number of practical information quality assessment and management methodologies, most significantly total data quality management (TDQM) (Huang et al., 1999; Wang, 1998), "assessment and improvement (AIM) quality" (AIMQ) (Lee et al., 2002), and data quality assessment (DQA) (Pipino et al., 2002). Subsequent research on these methodologies has found that they work very well in identifying and solving information quality issues, and that the underlying framework (i.e., Wang & Strong's quality dimensions) is robust and applicable to real-life information

quality situations. For example, Martin Eppler (2006) evaluated seven different information quality frameworks in the literature, and he determined that Wang and Strong's framework "is the only framework in the series of seven that strikes a balance between theoretical consistency and practical applicability" (p. 54).

Similarly, Carlo Batini et al. (2009) did a "systematic and comparative" review of 13 of the most well-known and established information quality methodologies (including TDQM, AIMQ, and DQA). They grouped them into four types (audit, complete, operational, and economic), based on how well they supported information quality assessment and improvement, as well as how they addressed technical and economic issues. They found that "audit methodologies [such as AIMQ and DQA] are more accurate than both complete and operational methodologies in the assessment phase… they are more detailed as to how to select appropriate assessment techniques… they identify all types of issues, irrespective of the improvement techniques that can or should be applied… the AIMQ methodology is the only information quality methodology focusing on benchmarking, that is, an objective and domain-independent technique for quality evaluation" (p. 28 and p. 38). As for operational methodologies, they found that

> one of the main contributions is the identification of a set of relevant dimensions to improve, and the description of a few straightforward methods to assess them. For example, TDQM is a general-purpose methodology and suggests a complete set of relevant dimensions and improvement methods that can be applied in different contexts … TDQM is comprehensive also from an implementation perspective, as it provides guidelines as to how to apply the methodology. (p. 28 and p. 35)

The main strength of Wang and Strong's (1996) information quality framework is that it covers all aspects of information quality—both objective (that is, "meets requirements") and subjective ("meets expectations"), as defined by Eppler (2006). Information quality must be measured along multiple dimensions—some objective (that is, that can be measured using objective methods) and some subjective (that is, that can only be measured based on how the user feels about it); any methodology that ignores one or the other will not provide a complete picture. It is certainly possible for information to have high objective quality and low subjective quality (Ge & Helfert, 2007). Stephanie Watts et al. (2009) emphasized that information quality must always be assessed "in use" because "information that is valuable and informative for one person may not be valuable and informative to the next, even when the information is objectively accurate and consistent" (p. 209). Extending the work done by Mouzhi Ge and Markus Helfert (2007), Phillip Woodall et al. (2014) stated that information quality issues can be grouped into four different quadrants based on a two-by-two conceptual model: issues that are either from the information's or user's perspective, and issues that are either context-independent or context-dependent.

Wang and Strong's (1996) framework is uniquely suited for information quality assessment because it includes both quantifiable and objective "context-independent" dimensions (for example, how accurate or consistent the information is) as well as quantifiable and objective "context-dependent" dimensions (for example, how complete or relevant the information is), and then measures them both using user-dependent subjective assessments (Huang et al., 1999).

My Proposed DQ Definition

Based on the research that showed the efficacy of Wang and Strong's (1996) framework of information quality categories and dimensions, in my 2019 study, I decided to apply it to the field of DQ. Because the framework in Wang and Strong (1996) is really a framework for measuring information quality, and documentation is information (that is, information that is transformed into knowledge by readers in a particular context for a particular reason), I believed that there was a strong basis for attempting to use this framework to accurately measure how readers define DQ.

Based on 81 responses from a broad, worldwide range of technical documentation readers from different fields (who were contacted via customer service personnel from various companies), I determined the single most important information quality dimension per ICRA category. I then created a reader-oriented DQ definition based on them. According to readers, high-quality documentation must be:

- Accurate (Intrinsic quality category)
- Relevant (Contextual quality category)
- Easy to Understand (Representational quality category)
- Accessible (Accessibility quality category)

Although this result might seem self-evident, I stated in my 2019 study that it provided a strong empirical underpinning for my claim that DQ could be defined using a narrow yet comprehensive set of clear and unambiguous information quality dimensions.

I then used this definition of DQ from the readers' point of view to propose a model for collecting meaningful and actionable feedback that could improve DQ and increase reader satisfaction. I claimed that my definition of DQ could also be used to provide reliable methods and metrics for measuring DQ, establish editing best practices, create a common DQ terminology, and help writers understand what is important to readers when feedback is unavailable (Strimling, 2018; 2021).

The framework proposed in my 2019 study seems to offer a good starting point for technical communicators who want to evaluate its practical applications, especially feedback and metrics (e.g., Johnson, 2021; Klein, 2024; Masycheff, 2023; Mui & Shwer, 2022; Yong, 2024). But to confirm its validity, it is

important to ensure that its underlying reader-derived definition of DQ is robust and truly representative of what readers want from high-quality documentation.

To do this, I ran a pilot study applying the Kano Model of customer satisfaction to the ICRA information quality categories and dimensions presented by Wang and Strong (1996). The Kano Model is used in many industries for decision analysis during the product/service development and design phases to "hear the voice of the customer" and determine which proposed features and customer requirements will have the greatest effect on customer satisfaction. The Kano Model has been successfully implemented and empirically tested in numerous studies of customer satisfaction across various product/service settings (for example, television sets and decorative table clocks (Kano et al., 1984); kindergartens, tourist destinations, mobile phones, and sports equipment (Mikulić, 2007); mobile banking, websites, and home appliances (Löfgren & Witell, 2008); pizzerias and video rental stores (Tontini et al., 2013); bicycles (Lin et al., 2017).

The Kano Model is also used in the field of UX to design strategies that help developers "hear the voice of the user" and focus only on the features that will have the most impact and present the biggest design opportunities (Spool, 2013; 2015; 2018). It is an indispensable tool for understanding user preferences (both stated and implied) precisely because it enables us to identify how they prioritize the possible features; organizations such as IBM, GitLab, the Nielson Norman Group, and the Baymard Institute use the Kano Model for UX improvements projects and feature prioritization (DeSanto, 2025; Gibbons, 2021; Holst, 2012; Olsen-Landis, 2017).

Kuan-Tsae Huang et al. (1999) make a strong case for considering information to be a "product" that has "features" that can increase or decrease customer satisfaction just like any other product. Bad information can have negative business impacts in the same way that other bad products do, and like other products a company makes, must be managed in the same way—by having a thorough knowledge of the consumers' needs and quality criteria. If this is not how a company regards its information, it will be impossible to deliver high-quality information consistently and reliably. If information is indeed a "product," then the Kano Model should certainly apply to it as well.

Based on this, I felt that the Kano Model would be well suited for measuring reader preferences (that is, "the voice of the reader") for the various DQ dimensions (which can be seen as "features" of the documentation, that is, the "information product"), and help build a stronger DQ definition.

Using The Kano Model Of Customer Satisfaction to Strengthen the DQ Definition

What Is the Kano Model of Customer Satisfaction?

In 1984, Noriaki Kano et al. proposed a model that stated that customer satisfaction/dissatisfaction with a product/service's quality attributes (that is, the features it

has) depends on their "state of physical fulfillment." Like Wang and Strong's (1996) framework, the Kano Model looks at both the objective aspect of quality (how much or how little the feature is fulfilled in the product/service) as well as the subjective aspect of quality (how customers feel about the feature) and then correlates the two.

The definition of the term "fulfillment" originally used by Kano et al. (1984) is somewhat vague. Charles Berger et al. (1993), in their significant and comprehensive review dedicated to the instruction, experience, ideas, and theories that have evolved from using Kano's methods, define the term fulfillment as a measure of functionality (i.e., how well or how badly the feature fulfills the task it is supposed to do). If the feature has its full functionality, then it can be implemented well; if it does not, then its implementation is not as good. Similarly, Josip Mikulić and Darko Prebežac (2011), in their extensive review of the most commonly used approaches to the classification of quality attributes according to the Kano Model, conclude that the correct way to look at fulfillment is to define it as "the provision (or non-provision) of the *benefits to be expected through the provision of the attribute* rather than the provision of the *attribute itself*" (p. 48, emphasis in original). In other words, we are asking customers to rate their satisfaction/dissatisfaction with the presence/absence of the feature's optimal implementation level and the performance of its benefits when provided.

Berger et al. (1993) also revised and improved the validity and reliability of the Kano Model in practice by making a number of modifications to it (e.g., rewording the allowed possible answers to the questions, recalibrating the evaluation table, adding a perceived importance rating (based on Hauser, 1991), assigning numerical values to responses, suggesting the use of continuous rather than discrete measurements, and so on). These improvements have been operationalized into a comprehensive and easy-to-use Kano Model Excel tool by Daniel Zacarias (2015) that enables UX and product management researchers to collect and analyze data about how proposed features will affect customer satisfaction.

Subsequent research on the Kano Model in the decades since its introduction (Löfgren & Witell, 2008; Madzík, 2018; Mikulić, 2007; Mikulić & Prebežac, 2011; Witell et al., 2013) has focused on attempts to improve its ease of use and the accuracy of its methodology, as well as comparing it to other customer satisfaction models. While a number of improvements have been suggested and the advantages of other models have been proposed, in general, researchers have found that the Kano Model (especially using continuous rather than discrete measurements and importance ratings, as per Berger et al. (1993) and implemented by Zacarias's(2015) tool) is a very valid and reliable method for determining which features will have the greatest impact on customer satisfaction.

Determining Feature Types

The Kano Model is based on two axes: customer satisfaction and functionality, as shown in Figures 4.1 and 4.2

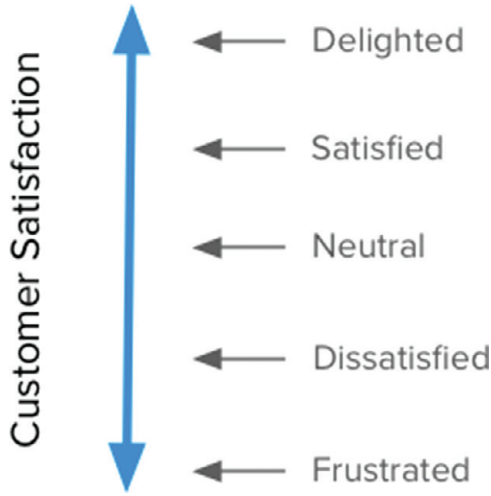

Figure 4.1. Kano Model customer satisfaction axis (from Zacarias, 2015)

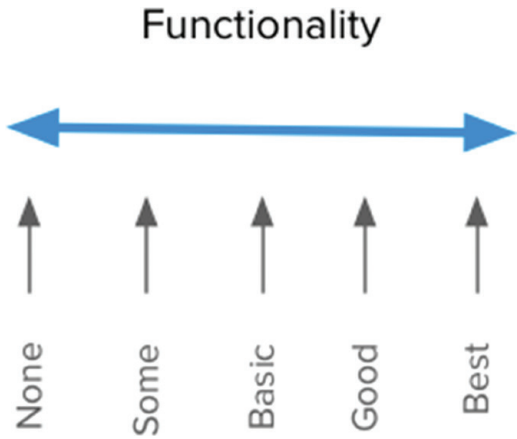

Figure 4.2. Kano Model functionality axis (from Zacarias, 2015)

The axes combine to form the following feature types, as shown in Figure 4.3.

- **Must-Be (M)**: These are features that are expected and taken for granted. They must be implemented properly; if they are not, customers will consider the product to be incomplete or bad. Their presence does not increase satisfaction or delight, but their absence decreases it. No matter how well they are implemented, customers will never be more than neutral about them; but if they are done badly, customers will quickly become frustrated.
- **One-Dimensional (O)**: These are features that have a one-to-one correlation between the level of functionality and customer satisfaction—the

better they are implemented, the greater the satisfaction; their presence delights customers and their absence frustrates them.

- **Attractive (A)**: These are features that the customer does not expect to be present, but when they are, they cause excitement and delight (even if they are not implemented as well as they could be). Their absence has very little effect on customer satisfaction, because customers were not expecting them in the first place (and therefore cannot be frustrated by their absence).
- **Indifferent (I)**: These features have no effect on customer satisfaction, regardless of how well they are implemented.

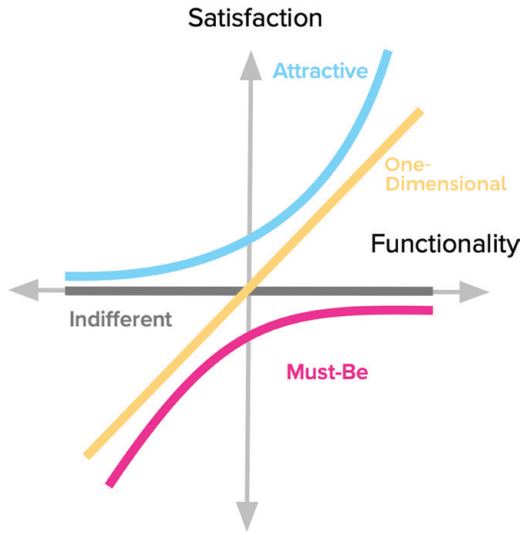

Figure 4.3. Kano Model feature types (based on Zacarias, 2015)

The following description of the Kano Model methodology is based on the modifications and improvements made by Berger et al. (1993) and operationalized by Zacarias (2015).

To determine a feature's type, customers are asked two types of questions about it:

- **Functional**: Determines how the customer feels when the feature is present (that is, well implemented) in the product
- **Dysfunctional**: Determines how the customer feels when the feature is absent (that is, not well implemented) from the product

The Functional/Dysfunctional questions have five allowed possible answers:

- **I Like It**: I would like it; enjoy it; it would be helpful to me.
- **I Expect It**: Must be that way; it is a basic need.
- **I Don't Care**: Neutral; wouldn't concern me; don't care.

- **I Can Live with It**: Dislike it but can live with it; inconvenience.
- **I Dislike It**: Extreme dislike; can't accept it; major issue.

Each answer is assigned a numerical value, which is plotted on a Functional/Dysfunctional two-dimensional grid, as shown in Figure 4.4. The value combinations create quadrants that are associated with the four feature types (Must-Be, One-Dimensional, Attractive, and Indifferent), as well as two other types:

- **Questionable (Q):** These are features whose type is unclear because the answers collected do not make sense. This could be because the question was not worded properly, or the respondents misunderstood what was being asked; for example, they like it when the feature is both present and absent.
- **Reverse (R):** These are features that respondents do not want in the product, and that need to be avoided; for example, they like it when the feature is not implemented and dislike it when it is.

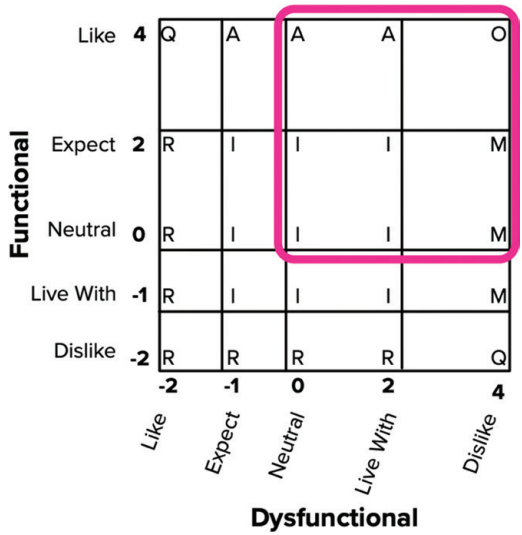

Figure 4.4. Kano Model rating two-dimensional scoring grid (from Zacarias, 2015)

Based on Figure 4.4, the rule of thumb for determining a feature's type can be summarized as follows:

- If the Functional value is *high* and the Dysfunctional value is *low*, then the feature type is *Attractive* (that is, it has a high potential to delight users if implemented well, but little effect otherwise; the greater the difference, the stronger the potential).
- If the Functional value is *low* and the Dysfunctional value is *high*, then the feature type is *Must-Be* (that is, it has a high potential to frustrate users if implemented badly, but little effect otherwise; the greater the difference, the stronger the potential).

- If both the Functional and Dysfunctional values are *high*, then the feature type is *One-Dimensional* (that is, its potential to delight or frustrate users correlates strongly to its level of implementation; the higher the values, the stronger the potential).
- If both the Functional and Dysfunctional values are *low*, then the feature type is *Indifferent* (that is, users are not delighted or frustrated by its level of implementation).

In addition to the Functional/Dysfunctional questions, a question is also asked about how important it is to the customer that the feature be implemented, rated on a unipolar Likert scale. This enables a finer level of differentiation among features.

These three values (Functional, Dysfunctional, and Importance) are combined to determine the proposed feature's type and its potential effect on customer satisfaction.

Prioritizing Feature Types

The heart of the Kano Model is the differentiation between feature types, which enables product/service planners and designers to know which features to prioritize to increase customer satisfaction. Due to time and resource constraints, not all features can be implemented, so it is critical to know which will have the greatest impact on customer satisfaction so those can be focused on.

Logically, those features that are classified as "Must-Be" should be the ones that get the highest priority. But, as Berger et al. (1993) said, the Kano Model shows us that not all customer requirements are equal—"improving performance on a Must-Be customer requirement that is already at a satisfactory level is not productive when compared to improving performance on a One-Dimensional or Attractive customer requirement" (p. 7). Improving Must-Be features does not increase customer satisfaction—they only have a negative effect if they are implemented badly, but no positive effect if they are implemented well. The question designers and project managers must ask is "after I address the Must-Be features to a satisfactory level, which feature type do I focus on next—One-Dimensional or Attractive?" Berger et al. (1993) suggested that a general guideline might be to first ensure that all Must-Be features are fulfilled, then be competitive with market leaders on the One-Dimensional features, and then add a few Attractive features to differentiate the product/service from others that are available. This too is logical—both Must-Be and One-Dimensional features have high Dysfunctional values (see the summarized "rule of thumb" above), which means that both have a high potential for frustrating customers. The difference, of course, is that One-Dimensional features also have a high potential for *delighting* them; Attractive features do as well but have a low potential for frustrating customers.

These feature types do not exist in a vacuum. Gerson Tontini et al. (2013) found that badly implemented Must-Be features negatively affect customer

satisfaction with well-implemented One-Dimensional and Attractive features and, to a lesser extent, badly implemented One-Dimensional features negatively affect customer satisfaction with well-implemented Attractive features. In other words, if Must-Be features are not implemented well, it does not matter how well the other features are implemented—the delight customers feel cannot compensate for the frustration they feel.

The addition of the Importance measure can be used to further prioritize and order features within a single type. For example, if many users say that a particular One-Dimensional feature is more important to them than another One-Dimensional feature, a case could be made to include one and not the other. Designers and project managers need to weigh all of these possibilities carefully, and the Kano Model provides them with a tool to help them do that.

Methods

Like in my 2019 study, this pilot study focused on finding the representative information quality dimension for each of the four ICRA information quality categories presented by Wang and Strong (1996) (as they related to documentation).

Pretest Questionnaires

Because the Kano Model questionnaire can be confusing and hard to use if not presented correctly, it is very important to test it before sending it to customers (Berger et al., 1993; Löfgren & Witell, 2008; Madzík, 2018; Zacarias, 2015). A pretest was therefore run with a limited audience of technical documentation readers (n = 12). Because of the intentionally small sample size and the exploratory nature of the survey, no quantitative data was collected about the Dysfunctional, Functional, and Importance values, and no demographic data was collected about the participants. However, a good deal of qualitative data was collected about the survey experience and how to reduce the cognitive load associated with the process, which was implemented in the pilot study. For example, it was determined that:

- The information quality dimensions must be divided into their respective ICRA quality categories; participants commented that 15 dimensions were too many to rate at one time using the Kano Model methodology.
- The allowed Kano Model answers must be defined on each page of the survey; participants commented that they could not remember what each option meant.

Pilot Study Questionnaires

In the pilot study itself, a number of technical communication department managers from various companies (in different industries, for example, software

development, robotics, oil and gas) were contacted via LinkedIn, and they were asked to talk to their respective customer service personnel about sending the revised Kano Model questionnaire to their technical documentation readers.

As in the pretest, no demographic data was collected about the participants. While standard Kano Model practice does include the collection of demographic data (for example, company and personal characteristics, familiarity with the product, use of competitors' products; see Berger et al., 1993), I chose to omit this step in the pilot study because I wanted to capture a wide spectrum of reader experiences with documentation across industries, rather than focus on demographic diversity. My primary goal was to determine if the Kano Model's approach to measuring customer satisfaction could be useful in improving DQ across a broad range of contexts, without introducing the additional complexity of segmenting respondents based on demographic factors.

Because I was trying to see if the Kano Model could help make the model I proposed in my 2019 study more robust, I felt that introducing demographic data would have added variables beyond this scope and potentially complicate the analysis. The limitation of this approach is that it required assuming a generic reader, as I was unable to focus on specific types of documentation readers (for example, readers with varying levels of experience, age groups, or other demographic distinctions). Nevertheless, because this was a pilot study, I believed that gathering insights from a broad range of readers across different industries would still provide valuable preliminary data.

Readers were asked to answer the Kano Model's Functional/Dysfunctional questions for each of Wang and Strong's (1996) 15 quality dimensions, as well as rate their Importance on a five-point Likert scale (1 = not important at all to 5 = extremely important). The questionnaire can be seen at https://www.surveymonkey.com/r/JSK9ZL9; note that the quality dimensions here are adjectives that describe the information in the documentation, and readers are being asked to consider the presence/absence of their best implementation level.

Data Analysis

A total of 47 readers responded to the pilot questionnaire (but only 43 of them rated all of the information quality dimensions). The Dysfunctional, Functional, and Importance values were entered into the Kano Model Excel tool provided by Zacarias (2015), which assigned a preliminary Kano Model feature type for each entry. As per Berger et al. (1993) and Peter Madzík (2018), information quality dimensions that received a Questionable rating from respondents (19 response entries out of 705) were removed from the analysis and were indicated by missing values.

The Kano Model Excel tool then calculated the mean weight per information quality dimension for all respondents and plotted them on the Functional/Dysfunctional two-dimensional grid, as shown in Figure 4.5.

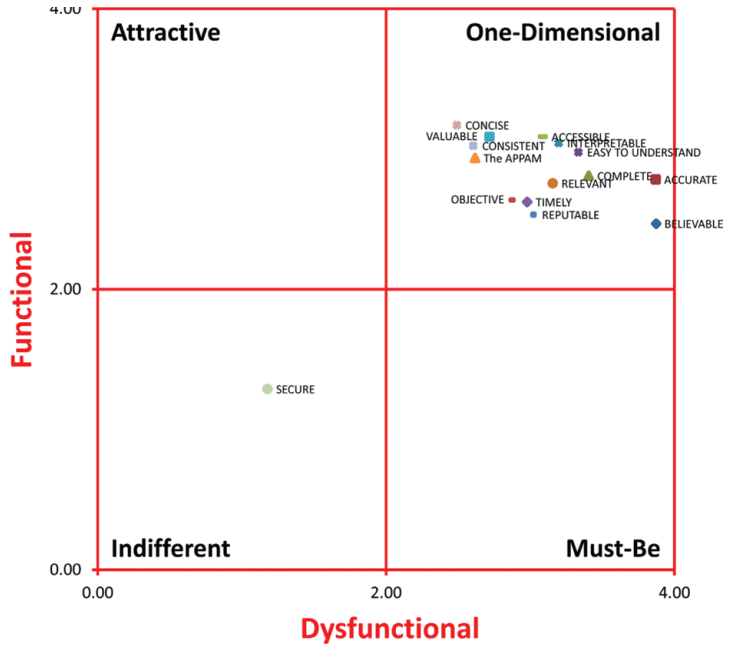

Figure 4.5. Functional/Dysfunctional two-dimensional scoring grid results (all feature types)

Each information quality dimension was assigned a Kano Model feature type based on their location on the grid. This was done to determine the relative priorities of each information quality dimension, as per the prioritization guidelines described previously. As can be seen in Figure 4.5, all but one of the information quality dimensions were located in the One-Dimensional quadrant; only Secure is in the Indifferent quadrant.

To enable better Kano Model data visualization within the One-Dimensional quadrant, I used "stack rankings" (modified from Moorman, 2012), which stacks the mean weights and standard deviations for the Dysfunctional, Functional, and Importance values of each information quality dimension. This enabled me to "zoom in" and determine which information quality dimensions had the highest potential for customer satisfaction. Based on the summarized "rule of thumb" described previously for determining a feature's type and the rationale for their priorities, the process was as follows (as suggested by Zacarias, 2018):

1. First, I looked for information quality dimensions whose Dysfunctional values were significantly higher than their Functional values (that is, their potential for frustration was greater than their potential for delight, almost like a Must-Be feature).

   To fine-tune the results, these dimensions were then sorted by Importance.

2. After this, I looked for information quality dimensions that had Dysfunctional and Functional values similar to but significantly higher than other dimensions (that is, a potential for frustration similar to their potential for delight, which would make them a purely One-Dimensional feature).

   To fine-tune the results, these dimensions were then sorted by Importance.

3. Lastly, I looked for information quality dimensions whose Dysfunctional values were significantly lower than their Functional values (that is, their potential for frustration was lower than their potential for delight, almost like an Attractive feature).

   To fine-tune the results, these dimensions were then sorted by Importance.

This process enabled me to identify the dimension that had the strongest combination of Dysfunctional, Functional, and Importance values per ICRA category, which I considered as representing the entire category.

It is important to note here that the Dysfunctional, Functional, and Importance values cannot be statistically compared to each other—the Dysfunctional and Functional values are rated on a -2 to 4 scale, and the Importance values are rated on a 1 to 5 scale. However, the Dysfunctional and Functional values can be compared to each other within and across information quality dimensions, and the Importance values can be compared to each other across information quality dimensions.

To determine if the differences in mean weights between the Dysfunctional and Functional values within and across the dimensions in each category (as well as the Importance values across them) were significant (set as $p < 0.05$), a one-way ANOVA test was run.

Results

Functional/Dysfunctional Two-Dimensional Grid Scoring

As mentioned previously, all but one of the information quality dimensions are located in the One-Dimensional quadrant; only Secure is in the Indifferent quadrant. One-Dimensional features have a one-to-one correlation between the level of functionality and customer satisfaction. In other words, most of the information quality dimensions are things that readers expect and that have a direct, proportional impact on reader satisfaction. For a discussion of this result, see the *Must-Be/Attractive Features vs. One-Dimensional Features* section.

Intrinsic Quality Category Results

Figure 4.6 shows the stack ranking results of the Dysfunctional, Functional, and Importance values for the Intrinsic information quality dimensions. The full

range of descriptive statistics is presented in Table 4.2, and the list of statistically significant differences is presented in Table 4.3.

Looking at these results, a number of interesting points appear:

- Readers have a statistically stronger *negative* reaction (that is, are the most frustrated, to use the Kano Model terminology) when the information in a document is not *Accurate* than a positive reaction (that is, delighted) when it is.
- They have a statistically stronger *negative* reaction when the information in a document is not *Believable* than a positive reaction when it is.
- They have the strongest *negative* reactions when the information in the documentation is not *Accurate* or *Believable*; these differences are statistically significant compared to both the *Objective* and *Reputable* dimensions.
- They have the strongest *positive* reaction when the information is *Accurate*; however, none of the differences are statistically significant.
- They think that the most important Intrinsic information quality dimension is *Accurate*; this difference is statistically significant only compared to the *Objective* and *Reputable* dimensions, but not to the *Believable* dimension.

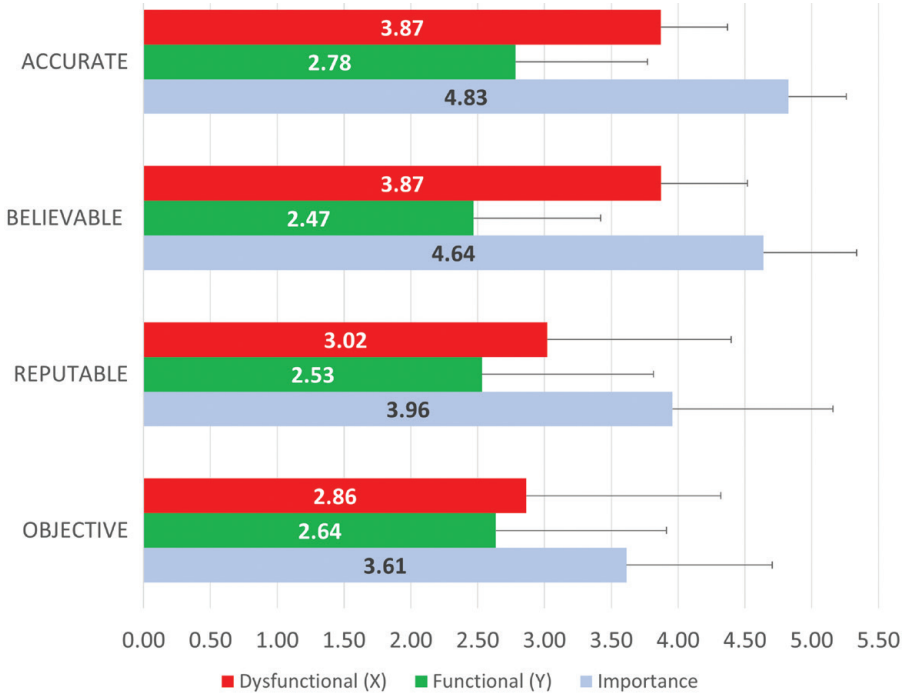

Figure 4.6. Intrinsic documentation quality dimension stack rankings

Table 4.2. Descriptive Statistics for Intrinsic Documentation Quality Dimensions

Quality Dimension	n	Dys Mean	Dys SD	Func Mean	Func SD	Imp Mean	Imp SD
Accurate	46	3.87	0.50	2.78	0.99	4.83	0.43
Believable	47	3.87	0.65	2.47	0.95	4.64	0.70
Objective	44	2.86	1.46	2.64	1.28	3.61	1.09
Reputable	47	3.02	1.38	2.53	1.28	3.96	1.20

Table 4.3. Statistically Significant Comparisons for Intrinsic Documentation Quality Dimensions

Quality Dimensions	*p* value
D_ACCURATE vs F_ACCURATE	< 0.0000
D_BELIEVABLE vs F_BELIEVABLE	< 0.0000
I_ACCURATE vs I_OBJECTIVE	< 0.0000
I_BELIEVABLE vs I_OBJECTIVE	< 0.0000
I_ACCURATE vs I_REPUTABLE	< 0.0000
D_ACCURATE vs D_OBJECTIVE	0.0001
D_BELIEVABLE vs D_OBJECTIVE	0.0001
D_BELIEVABLE vs D_REPUTABLE	0.0010
D_ACCURATE vs D_REPUTABLE	0.0011
I_BELIEVABLE vs I_REPUTABLE	0.0021

(D_ = Dysfunctional, F_ = Functional, I_ = Importance)

The combination of these results suggests that either *Accurate* or *Believable* might be the Intrinsic information quality dimension that readers want us to focus on to ensure high-quality documentation; in my 2019 study, the representative Intrinsic information quality dimension was *Accurate*. For a discussion of this result, refer to the *Documentation Accuracy vs. Documentation Believability* section.

Contextual Quality Category Results

Figure 4.7 shows the stack ranking results for the Dysfunctional, Functional, and Importance values for the Contextual information quality dimensions. The full range of descriptive statistics is presented in Table 4.4 and the list of statistically significant differences is presented in Table 4.5.

Looking at these results, a number of interesting points appear:

- Readers have a statistically stronger *negative* reaction (that is, are the most frustrated) when the information in a document is not *Complete* than a positive reaction (that is, delighted) when it is.
- They have the strongest *negative* reaction when the information in the documentation is not *Complete;* this difference is statistically significant only compared to the *Appropriate Amount* dimension.
- They have the strongest *positive* reaction when the information is *Valuable*; however, none of the differences are statistically significant.
- They think that the most important Contextual information quality dimension is *Complete*; this difference is statistically significant only compared to the *Appropriate Amount* and *Valuable* dimensions.

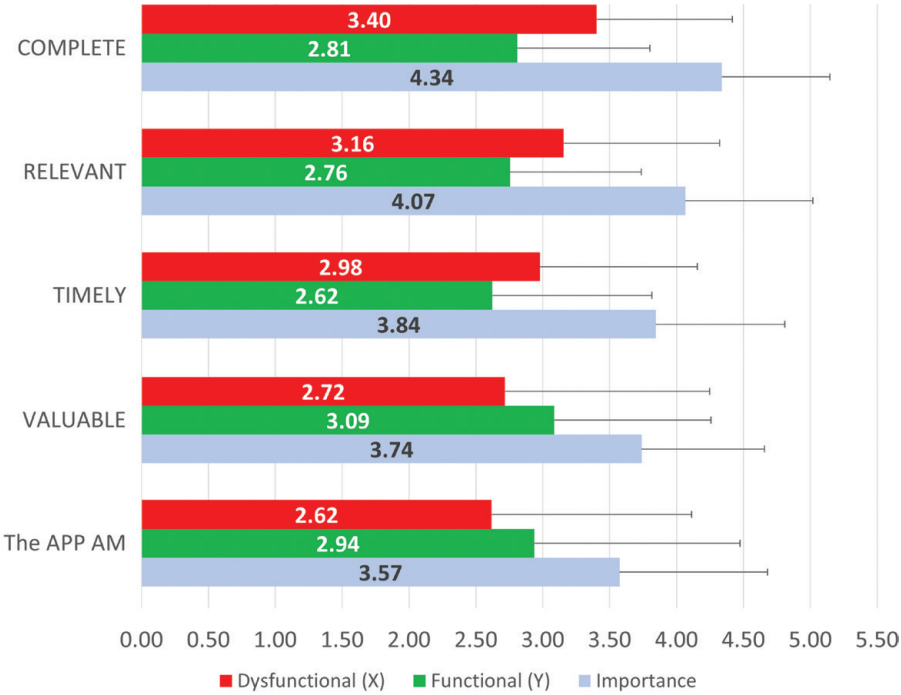

Figure 4.7. Contextual documentation quality dimension stack rankings

Table 4.4. Descriptive Statistics for Contextual Documentation Quality Dimensions

Quality Dimension	n	Dys Mean	Dys SD	Func Mean	Func SD	Imp Mean	Imp SD
Complete	47	3.40	1.01	2.81	0.99	4.34	0.81

Quality Dimension	n	Dys Mean	Dys SD	Func Mean	Func SD	Imp Mean	Imp SD
Relevant	45	3.16	1.17	2.76	0.98	4.07	0.95
The Appropriate Amount	47	2.62	1.50	2.94	1.54	3.57	1.11
Timely	45	2.98	1.18	2.62	1.19	3.84	0.97
Valuable	46	2.72	1.53	3.09	1.17	3.74	0.92

Table 4.5. Statistically Significant Comparisons for Contextual Documentation Quality Dimensions

Quality Dimensions	*p* value
D_COMPLETE vs F_COMPLETE	0.0050
I_COMPLETE vs I_APPAM	0.0012
D_COMPLETE vs D_APPAM	0.0287
I_COMPLETE vs I_VALUABLE	0.0222

(D_ = Dysfunctional, F_ = Functional, I_ = Importance)

The combination of these results suggests that *Complete* might be the Contextual information quality dimension that readers want us to focus on to ensure high-quality documentation. This is different from the Contextual quality category result I reported in my 2019 study; there, the representative information quality dimension was *Relevant*. For a discussion of this result, refer to the Documentation Completeness vs. Documentation Relevance section.

Representational Quality Category Results

Figure 4.8 shows the stack ranking results for the Dysfunctional, Functional, and Importance values for the Representational information quality dimensions. The full range of descriptive statistics is presented in Table 4.6, and the list of statistically significant differences is presented in Table 4.7.

Looking at these results, a number of interesting points appear:

- Readers have a statistically stronger *positive* reaction (that is, are the most delighted) when the information in a document is *Concise* than a negative reaction (that is, frustrated) when it is not.
- They have the strongest *negative* reaction when the information in the documentation is not *Easy to Understand*; this difference is statistically

significant only compared to the *Concise* and *Consistent* dimensions.

- They have the strongest *positive* reaction when the information is *Concise*; however, none of the differences are statistically significant.
- They think that the most important Representational information quality dimension is *Easy to Understand*; however, none of the differences are statistically significant.

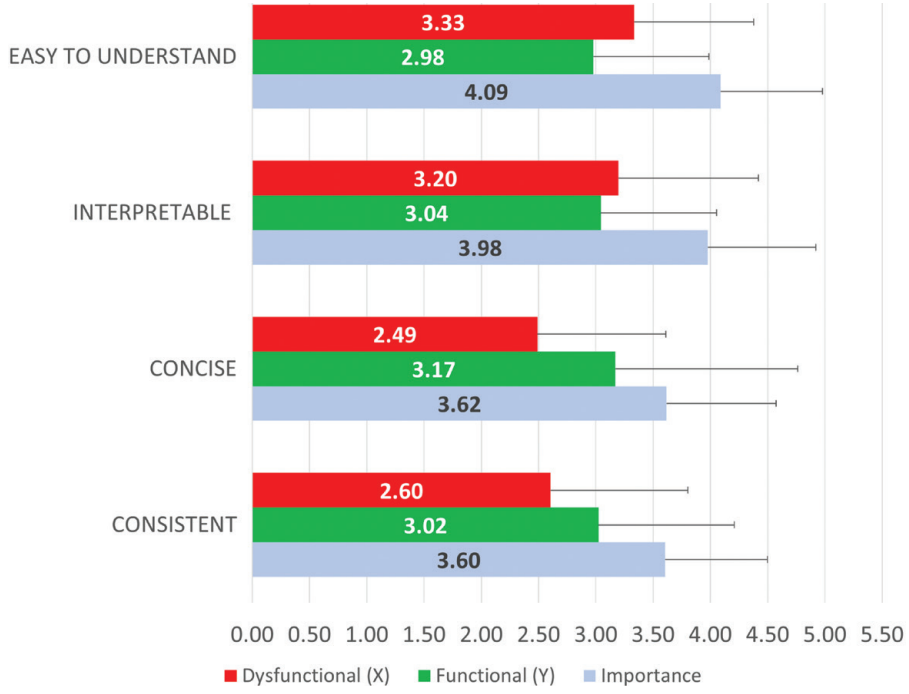

Figure 4.8. Representational documentation quality dimension stack rankings

Table 4.6. Descriptive Statistics for Representational Documentation Quality Dimensions

Quality Dimension	n	Dys Mean	Dys SD	Func Mean	Func SD	Imp Mean	Imp SD
CONCISE	47	2.49	1.12	3.17	1.59	3.62	0.96
CONSISTENT	43	2.60	1.20	3.02	1.18	3.60	0.89
EASY TO UNDERSTAND	45	3.33	1.04	2.98	1.01	4.09	0.89
INTERPRETABLE	46	3.20	1.22	3.04	1.01	3.98	0.94

Table 4.7. Statistically significant comparisons for Representational Documentation Quality Dimensions

Quality Dimensions	*p* value
D_EASYTOUNDERSTAND vs D_CONCISE	0.0030
D_EASYTOUNDERSTAND vs D_CONSISTENT	0.0174
D_INTERPRETABLE vs D_CONCISE	0.0179
F_CONCISE vs D_CONCISE	0.0185

(D_ = Dysfunctional, F_ = Functional)

The combination of these results suggests that *Easy to Understand* might be the Representational information quality dimension that readers want us to focus on to ensure high-quality documentation, which is similar to the representative Representational information quality dimension I reported in my 2019 study.

It is interesting to note here that the *Concise* dimension is the only dimension (in any quality category) whose Functional value is significantly higher than its Dysfunctional value. This implies that documentation readers are very delighted with documentation that is concise but are not overly frustrated if it is not. Quite the contrary—readers are more frustrated when a document is not easy to understand than when it is not concise. For a discussion of this result, refer to the *Documentation Conciseness vs, Documentation Understandability* section.

Accessibility Quality Category Results

Figure 4.9 shows the stack ranking results for the Dysfunctional, Functional, and Importance values for the Accessibility information quality dimensions. The full range of descriptive statistics is presented in Table 4.8, and the list of statistically significant differences is presented in Table 4.9.

Looking at these results, a number of interesting points appear:

- Readers have the strongest *negative* reaction (that is, are the most frustrated) when the information in the documentation is not *Accessible*.
- They have the strongest *positive* reaction (that is, are the most delighted) when the information is *Accessible*.
- Readers think that the most important Accessibility information quality dimension is *Accessible*.

All of these differences are statistically significant.

The combination of these results indicates that *Accessible* is the Accessibility information quality dimension that readers want us to focus on to ensure high-quality documentation, which is similar to the representative Accessibility information quality dimension I reported in my 2019 study.

As opposed to readers' very strong positive feelings about the *Accessible* dimension, the *Secure* information quality dimension is the only one that is classified as an Indifferent feature type (as shown in Figure 4.5). For a discussion of this result, refer to the *Documentation Accessibility vs. Documentation* section.

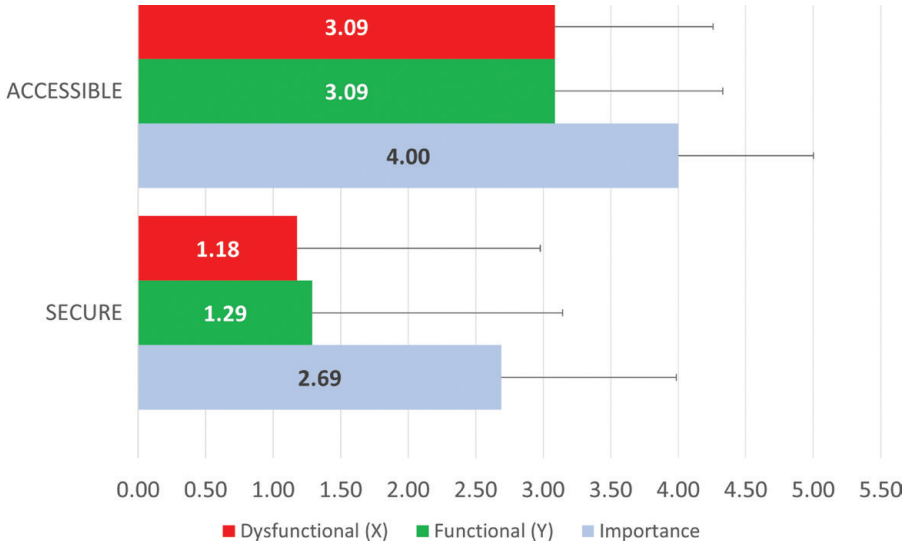

Figure 4.9. Accessibility documentation quality dimension stacked rankings

Table 4.8. Descriptive Statistics for Accessibility Documentation Quality Dimensions

Quality Dimension	n	Dys Mean	Dys SD	Func Mean	Func SD	Imp Mean	Imp SD
Accessible	46	3.09	1.17	3.09	1.24	4.00	1.00
Secure	45	1.18	1.80	1.29	1.85	2.69	1.30

Table 4.9. Statistically Significant Comparisons for Accessibility Documentation Quality Dimensions

Quality Dimensions	*p* value
D_ACCESSIBLE vs D_SECURE	< 0.0000
F_ACCESSIBLE vs F_SECURE	< 0.0000
I_ACCESSIBLE vs I_SECURE	< 0.0000

(D_ = Dysfunctional, F_ = Functional, I_ = Importance)

Discussion

It is important to note that I cannot clearly state (as I did in my 2019 study) which dimension best represents each quality category based on statistical significance (even though many of the within- and across-dimension differences are statistically significant). However, the results still indicate trends in the data that we can use to analyze the DQ definition I proposed there.

Must-Be/Attractive Features Versus One-Dimensional Features

As mentioned previously, almost all the information quality dimensions are classified as One-Dimensional features (except for *Secure*, which is an Indifferent feature), that is, they have a direct, proportional impact on reader satisfaction. But what does it mean that there are no Must-Be features?

I believe that this is a very encouraging result. Must-Be features do not increase customer satisfaction; they can only decrease it. Because these DQ dimensions are One-Dimensional features, improving their implementation in our documentation will have a tangible effect on how readers see the quality of our documentation. These are things that we should invest time and effort in to ensure that they are as well implemented as possible—the return on investment will be substantial. The better we do these features, the more delighted our readers will be. However, because we only have limited time and resources, and cannot improve all of them all at once, we still need to focus our efforts on those features that will have the greatest impact on reader satisfaction to cover all aspects of DQ (that is, that represent each of the ICRA categories). The fact that some of these One-Dimensional dimensions do have characteristics of Must-Be features (for example, the *Accurate* and *Complete* dimensions, whose Dysfunctional values are significantly greater than their Functional values) makes this easier.

But what about the lack of Attractive features?

This result should act as a warning to us as technical communicators. Attractive features, while increasing customer satisfaction, are, by definition, things that customers do not expect. These DQ dimensions are One-Dimensional features, which means that they are expected by readers and cannot be ignored. Even if we do not or cannot implement all of them now, readers still want them to be implemented well in the documentation they use and will be frustrated if they are not—we must plan for them in future releases or when we have the time/resources. While they are not necessarily the representative dimensions of each information quality category (that is, how readers define DQ), they still play an important role in improving reader satisfaction. The fact that some of these One-Dimensional dimensions do have characteristics of Attractive features (for example, the *Concise* dimension, whose Functional value is significantly greater than its Dysfunctional value) can act as a guideline for prioritization.

Documentation Accuracy Versus Documentation Believability

What do readers mean when they say they want the information in the documentation to be accurate, and how does this differ from wanting it to be believable? Based on Wang and Strong (1996), these Intrinsic information quality dimensions are defined as follows:

- **Accurate:** The information in the documentation is correct, reliable, and certified free of error.
- **Believable:** The information in the documentation is true, real, and credible.

No statistical difference was found in this study between the combination of the Dysfunctional, Functional, and Importance values for the *Accurate* and *Believable* dimensions. Unlike in my 2019 study, I cannot clearly state which of these two One-Dimensional dimensions best represents the Intrinsic quality category—both have a higher potential for frustrating readers with their absence than delighting them with their presence (almost like Must-Be features).

However, the difference between the *Accurate* and *Believable* dimensions can still be meaningful. Information in a document can be accurate but not believable, or believable but not accurate. Readers might trust well-presented information, even though it is inaccurate, and they might dismiss accurate information because they do not trust the source (for example, because the information feels like it is presented in a haphazard way, or because they have had bad experiences with the accuracy of the information before (Falla, 2018; Snyder, 1998)). We see this frequently now with the advent of so-called "fake news"—we often accept/reject what we are told based on how we feel about the source, without fact checking it. Watts et al. (2009) make a further point about the difference between accurate information and believable information—accuracy can be measured objectively; believability is subjective: information that is believable to one reader (for example, a novice user) might be less believable to another (for example, an expert).

As an interesting thought experiment, I prompted OpenAI's ChatGPT-3.5 LLM with the statement "Act as a technical documentation user, and tell me which is better—accurate information or believable information," to which the ChatGPT-generated text responded

> As a technical documentation user, I would prioritize accurate information over believable information… Accurate information ensures that users can make informed decisions, troubleshoot problems, and understand the intricacies of a system or process correctly. It helps prevent errors, misunderstandings, and potential risks that may arise from following incorrect instructions or guidelines. On the other hand, believable information may not necessarily be accurate. While

> it might sound plausible, it could lead to incorrect conclusions, poor decisions, or wasted time and effort for users trying to implement the guidance provided (OpenAI, 2023).

In the case of technical communication, the accuracy of information is logically more important than its believability: inaccurate concept/task information is useless to readers, regardless of how believable it is. Without the ability to verify the objective accuracy of the information, all readers have to go on is their subjective feelings of belief; but even so, it is their belief in its accuracy.

In the information quality literature, the *Accurate* dimension is one of a small set of dimensions that almost all frameworks and methodologies include; the *Believable* dimension is hardly ever included. For example, Mouzhi Ge et al. (2011) found that the *Accurate* dimension was included in all of them; the *Believable* dimension was only included in three. Similarly, Corinna Cichy and Stefan Rass (2019) reviewed and compared 12 general-purpose information quality methodologies that contained information quality definitions and assessment and improvement processes and found that the *Accurate* dimension was included in eight of them; again, the *Believable* dimension was only included in three. Clearly, information accuracy is a more critical part of quality than information believability. Wang and Strong (1996) purposely titled their paper "Beyond Accuracy" because they felt that data quality improvement efforts tended to focus too narrowly on accuracy and ignored any other dimensions.

Based on this, it appears logical to state (as I did in my 2019 study) that *Accurate* is the information quality dimension that best represents the Intrinsic quality category and therefore represents how readers define DQ.

Documentation Completeness Versus Documentation Relevance

What do readers mean when they say they want the information in the documentation to be complete? Why does this differ from the result in my 2019 study that stated that documentation relevance was more important? Based on Wang and Strong (1996), these Contextual information quality dimensions are defined as follows:

- **Complete:** The information in the documentation is of sufficient breadth, depth, and scope for the task at hand.
- **Relevant:** The information in the documentation is applicable and helpful for the task at hand.

No statistical difference was found in this study between the combination of the Dysfunctional, Functional, and Importance values for the *Complete* and *Relevant* dimensions. In my 2019 study, there was also no statistical significance between these two Contextual information quality dimensions. Even so, I claimed there that it was logical to assume that the *Relevant* dimension best represented

the Contextual quality category. Because documentation is never read in a vacuum and is only used in context, the usability of its information depends mainly on its ability to help readers do the "task at hand", which can be accomplished even if there are issues with how complete the information is. On the other hand, irrelevant information that is complete is still irrelevant.

However, I can also make a counterclaim that, if a concept/task that is irrelevant to the reader is provided, it does not really matter if it is complete or not; but even incomplete information can still be "applicable and helpful" to the reader to some degree.

It is important to consider how we look at information completeness. Batini et al. (2009) found that in the 13 information quality methodologies they reviewed there was "substantial agreement" about how it should be defined, namely, is information missing or not. But how this is measured makes a difference—does this mean that completeness is a binary, black-and-white measurement, "yes, the information is there; no, it is not?" Watts et al. (2009) stated that completeness can be measured objectively, which implies that it is. Information relevance, on the other hand, they say is subjective, and depends on the reader: information that is relevant for one reader (for example, a novice user) might be less relevant to another reader (for example, an expert).

But in Wang and Strong's (1996) information quality framework, *Complete* is a Contextual information quality dimension, *not* an Intrinsic information quality dimension. As Yang Lee et al. (2002) explained, completeness is an Intrinsic dimension only when simply referring to missing information, but it is a Contextual dimension when referring to missing information needed by users for the "task at hand". Information can be complete only when all the relevant information is included; adding irrelevant information will not make a document more complete, quite the contrary (Carey et al., 2014).

In terms of DQ, this is a critical point—the completeness of the information provided to readers has a direct impact on whether they can do what they need to do or know what they need to know. We can make the same claim for the subjectiveness of information completeness that Watts et al. (2009) made for information relevance previously—it also depends on the reader: information that is complete for one reader (e.g., an expert user) might be less complete for another reader (e.g., a novice). John Carroll and Hans van der Meij (1996) put it very succinctly when they say that "completeness is always a matter of degree" (p. 73).

There is clearly a difference between the *Complete* and *Relevant* dimensions. Information in a document can be complete but irrelevant, or relevant but incomplete—which is worse?

- Readers might be knowledgeable enough to skip irrelevant information, but be unable to fill in missing concept/task information OR
- Readers might be experienced enough to fill in missing information but not be sure if the concept/task is relevant to them.

As we can see from the results of this Kano Model study, readers are more frustrated with incomplete information than they are delighted with complete information (almost like a Must-Be feature); the same cannot be said for irrelevant information. In my 2019 study, only the relative importance of the information quality dimensions was measured; here, we are looking at the negative and positive reactions readers have, as well as the potential impact they have on readers' frustration and delight. This extra layer of information enables us to better discriminate between these two Contextual information quality dimensions.

Based on this, it appears logical to state (not like I did in my 2019 study) that *Complete* is the information quality dimension that best represents the Contextual quality category and therefore represents how readers define DQ.

Documentation Conciseness vs. Documentation Understandability

What do readers mean when they say they would like the information in the documentation to be concise? Based on Wang and Strong (1996), this dimension is defined as follows:

- **Concise:** The information in the documentation is compactly represented without being overwhelming (that is, it is brief in presentation, yet complete and to the point).

More information is not necessarily a good thing and can present problems for readers who are trying to apply it and put it into practice. This is a key element of minimalism—not everything needs to be documented; there are some things that we assume readers already know or can figure out easily on their own. Minimalism does not mean writing fewer words, it means making your writing more concise and to the point, knowing what to include and what not to include so readers can focus only on the essentials (Carey et al., 2014; Carroll & van der Meij, 1996; Virtaluoto et al., 2021).

Information that is concise helps readers focus on knowing what they need to know or doing what they need to do. Readers are busy people—they do not usually have time to read through lots of documentation. Reading documentation is not a high priority for them and is often seen as a last resort. Making the information concise helps readers use their time more efficiently.

Compare this to the definition of the *Easy to Understand* information quality dimension, which was the representative Representational information quality dimension reported in my 2019 study:

- **Easy to Understand:** The information in the documentation is clear, without ambiguity, and easily comprehended.

Documentation that is easy to understand uses clear and unambiguous language, presents complicated information as tables or bulleted lists, uses visually

effective figures and concrete examples, avoids jargon, and so on. Concise information can do this too, of course, but it relies more on the knowledge level of the reader. Minimalism is, by design, very user-centered, and the prior knowledge and familiarity of the audience with what is being documented determines how the information is written and presented. Documentation that is simply easy to understand but not concise will enable better comprehension for a broader range of readers. While both conciseness and understandability are subjective measures that depend on the reader (similar to the level of documentation completeness discussed previously), easy-to-understand information will be helpful for a much wider audience. Information in a document can be easy to understand but not concise, or concise but not easy to understand. In the former case, the document is still usable; in the latter, its usability is seriously reduced.

As we can see from the results of this Kano Model study, readers are more delighted with information that is concise than they are frustrated with information that is not (almost like an Attractive feature). It is also clear that they are more frustrated with information that is not easy to understand than they are frustrated with information that is not concise. While readers would prefer the documentation to be concise, it is not a "make-or-break" issue for them—but they will not tolerate documentation that is not easy to understand.

Based on this, it appears logical to state (as I did in my 2019 study) that *Easy to Understand* is the information quality dimension that best represents the Representational quality category and therefore represents how readers define DQ.

Documentation Accessibility Versus Documentation Security

What do readers mean when they say they want the information in the documentation to be accessible, but do not care (that is, they are "indifferent", to use the Kano Model terminology) if it is secure? Based on Wang and Strong (1996), these Accessibility information quality dimensions are defined as follows:

- **Accessible:** The information in the documentation is available or easily and quickly retrievable.
- **Secure:** Access to the information in the documentation can be restricted, and hence, kept secure.

The difference between these dimensions is clear. Unlike the dimensions in the other information quality categories, these two dimensions are almost mutually exclusive. While information in a document can be accurate, believable, complete, relevant, easy to understand, and concise, it cannot be both accessible and secure at the same time. If access to the information can be restricted, then it cannot also be available or easily and quickly retrievable.

There might be many good reasons for the information in a document to be secure—intellectual property rights, proprietary data, or regulatory requirements. But, in general, these issues are not the readers' concerns, they are the concerns of

the company that created the documentation. As we can see from the results of this Kano Model study, readers do not care about this aspect of DQ.

What they do care about, however, is their ability to find and retrieve the information they need, easily and quickly. There are many best practices for this, for example, comprehensive indexes and tables of contents, well-organized document structures and information chunks (such as those used in topic-based modular documentation), clearly marked headings, optimized search algorithms, easy navigation through the information by reducing the number of required clicks, complete metadata, and so on (Carey et al., 2014; Cheung, 2016; Strimling & Corbin, 2009). All of these will increase information accessibility.

Information accessibility can also enable readers to personalize (or fix) the documentation themselves so it is more accurate, relevant/complete, easy to understand, or accessible—user-created, personalized content is very important to readers, and can help improve DQ (Bowman, 2022; Richardson, 2019).

Based on this, it appears clear to state (as I did in my 2019 study) that *Accessible* is the information quality dimension that best represents the Accessibility quality category and therefore represents how readers define DQ.

Practical Applications

The ultimate goal of this research is to propose concrete solutions that address some of the most important issues faced by technical communication practitioners, and answer the call from Andersen and Hackos (2018) to provide "research findings based on empirical evidence [that] can help practitioners make more informed decisions and stronger business cases for resources, initiatives, or changes needed to improve processes and solve problems" (p. 97). As stated at the very beginning of this chapter, collecting meaningful and actionable feedback from readers about the technical documentation they use and creating consistent and reliable DQ metrics that technical communicators can present to their managers are crucial aspects of improving DQ, and can only be done by understanding how readers themselves define DQ.

By using the focused, clearly defined, and reader-derived definition of DQ proposed in my 2019 study, and refining it using the results of the Kano Model pilot study presented here, I believe that we now have a very strong DQ definition that can be used to do exactly this.

Collecting Meaningful and Actionable Documentation Quality Feedback

In my 2019 study, I suggested a way to turn the readers' DQ definition into a usable feedback collection method, which could be applied to a number of techniques for testing documentation usability (for example, usability edits, surveys, and training classes). According to this, each of the four representative

information quality dimensions can be converted into a question; this will enable us to "focus only on the most important issues from the readers' point of view, ask the fewest possible number of questions that can cover all of these important issues, [and] use terminology that can be clearly and universally understood by all respondents" (Strimling, 2019. p. 9). Modifying my 2019 proposed DQ definition based on our Kano Model results, the questions we should ask will now be:

- Could you find the information you needed in the document?
- Was the information in the document accurate?
- Was the information in the document complete?
- Was the information in the document easy to understand?

The answers that readers provide to these questions will be unambiguous, easily understood, and help technical communicators make informed decisions about how to improve their documentation.

Providing Reliable Metrics for Measuring Documentation Quality

Similarly, I previously suggested that the readers' DQ definition could be turned into a way to "classify and sort existing internal or external feedback, [which] can then be presented to management as clear and reliable metrics about the documentation that will help determine where more emphasis might need to be invested" (Strimling, 2019, p. 23). Measuring how accurate, complete, easy to understand, and accessible readers perceive our documentation to be is essential for gauging its quality. The information quality dimensions reflected in our Kano Model results serve as key indicators of DQ and can offer valuable insights into areas where improvement might be needed. Metrics, as Spool (2018) says, are about knowing what to measure, knowing what the measurements mean, and about knowing how to use them to improve our readers' experience. By systematically evaluating DQ based on these dimensions, technical communicators can make stronger business cases for how and where to allocate their limited resources to maximize reader satisfaction, as well as demonstrate the impact that high-quality documentation can have.

Teaching Students about Documentation Quality

But aside from the effect that this empirically based and reader-derived definition of DQ can have on technical communication practitioners, there are also clear practical applications for academics here as well. By using this real-life DQ definition, we can bridge the gap between theory and practice and create evidence-based materials to teach the next generation of technical communicators how to write documentation that is helpful and valuable to readers. Moreover, incorporating real-world DQ feedback and metrics into academic curricula can provide students with tangible guidelines and best practices that they can use when they enter the workforce.

Conclusion and Future Directions

The stated purpose of this paper was to take the reader-derived definition of DQ_I previously proposed (Strimling, 2019) and make it more robust by applying the Kano Model of customer satisfaction to it. It was hoped that, by using the Kano Model's unique approach to classifying feature types and customer requirements, I could better determine which of the information quality dimensions presented in Wang and Strong (1996) really represented each information quality category, as it applies to documentation.

Overall, it seems to me that the application of the Kano Model's methodology to the reader-derived definition of DQ_I proposed in my 2019 study has indeed enhanced the DQ_definition. While not providing outright support for the use of the particular information quality dimensions I proposed there, especially for the Contextual information quality category (that is, emphasizing documentation completeness instead of relevance), nevertheless, there are strong similarities in the trends observed in both studies. Both underscore the value of documentation accuracy, ease of understanding, and accessibility to readers in convincing ways, and show that the context in which documentation is used plays an important role in DQ.

In my opinion, the differences between the results in my 2019 study and the results in this Kano Model pilot study can be attributed to the differences in their emphasis. The Wang and Strong (1996) framework I used there, with its structured and hierarchical approach to identifying dimensions of information quality, focuses on readers' ratings of the relative importance of these dimensions. This approach offers a detailed understanding of reader opinions about what they prioritize in the documentation and serves as a solid foundation for defining DQ.

However, DQ is more than just the importance of these dimensions to readers. Readers might consider many information quality dimensions "important", but that does not necessarily mean that focusing on these will lead to increased customer satisfaction. According to International Organization for Standardization (ISO) 9000:2015, the term "customer satisfaction" is defined as "[The] customer's perception of the degree to which the customer's expectations have been fulfilled" (n.p.). This definition, though, comes with the following important caveats:

- These expectations might not be known to the company, or even to the customer, until the product or service is delivered.
- It might be necessary to fulfill a customer's expectation even if it is not stated, generally implied, or obligatory (that is, a "requirement").
- Even when a customer's requirements are agreed upon and fulfilled, this does not necessarily ensure high customer satisfaction.

Rather than just simply measure the relative importance of the quality dimensions (like I did in my 2019 study), the Kano Model goes beyond this and

measures customer satisfaction by categorizing the information quality dimensions into delighters/frustraters, capturing both expected and unexpected aspects of DQ. This focus on measuring customer satisfaction adds another layer to our understanding of DQ, aligning closely with the ISO 9000:2015 definition of customer satisfaction.

According to Feng-Han Lin et al. (2017), there are many ways to understand and prioritize the Kano Model results, and its main strength is that it provides decision makers with guidelines that can help them decide which features might have the biggest impact on customer satisfaction (Berger et al., 1993; Zacharias, 2018). In other words, the Kano Model is an important tool to help us understand what impacts customer satisfaction with DQ, but we cannot consider its results in a vacuum.

In this pilot study, because no demographic data was collected, it was impossible to segment the types of documentation readers who answered the Kano Model survey, which certainly had an impact on the responses. The need for participant segmentation is clearly important and can provide useful insights. For example, Jan Moorman (2012) observed distinct reactions to product features based on users' market savvy. By segmenting responses by profile, such as early, late, and non-adopters, the analysis became much more focused. This kind of segmentation ensures that the analysis reflects the varied needs and perceptions of different user groups. Similarly, Kano et al. (1984) themselves found differences between how men, women, married, unmarried, younger, and older people rated features of television sets.

It is important to remember that "technical documentation readers" are not a monolithic entity—they have different experiences (e.g., they are advanced or novice users), they have different needs (e.g., they want to understand a concept or do a task), and they have different backgrounds (e.g., they do not speak English as a first language, or they are older users). However, because the goal of this pilot study was to see if the Kano Model's approach to measuring customer satisfaction might be useful in improving DQ, the focus here was on capturing general trends and insights that might be broadly applicable, rather than analyzing variations based on demographic characteristics.

Future DQ studies using the Kano Model should, of course, attempt to gather and incorporate detailed participant information and examine the influence of demographic variables on perceptions of DQ. This will enable segmentation of the responses based on relevant reader profiles, which will enhance the robustness and applicability of this study's conclusions, ensure that the findings are more precise and tailored to different reader groups, and contribute to a more comprehensive understanding of how different groups experience and evaluate documentation.

I believe that the combination of Wang and Strong's (1996) information quality framework and the Kano Model can lead to a better understanding of DQ from the reader's point of view. Each approach has its strengths and ultimately

complement each other in defining DQ in a clear, comprehensive, and empirically based manner. The broader perspective offered by the Kano Model allows us to go beyond merely looking at the importance of the dimensions to readers and explore their satisfaction levels regarding each dimension's presence or absence in the documentation. By integrating the results presented in my 2019 study based on Wang and Strong (1996), with the results suggested by the Kano Model here, we can get a much clearer picture of what our readers want from the documentation we send them and how they themselves define high-quality documentation—*accurate, complete (*or *relevant), easy to understand,* and *accessible.*

Based on the results of this pilot study, it seems to me that the Kano Model is a useful tool for defining how readers define DQ and determining what they want from the documentation we send them. Its focus on the positive and negative attitudes that readers have regarding certain information quality dimensions opens up a new avenue for research into DQ. This study should be seen as a beginning attempt at applying the Kano Model's well-founded methodology to the field of DQ, and be used as a starting point for building a framework for collecting meaningful and actionable feedback, creating consistent and reliable DQ metrics, and providing guidelines for use in academic technical communication courses to teach students about real-life reader-oriented DQ measures.

References

Andersen, Rebekka, & Hackos, JoAnn. (2018). Industry perspectives on academic research in technical communication. *CIDM Best Practices*, *20*(4), 91–98

Arazy, Ofer, Kopak, Rick, & Hadar, Irit. (2017). Heuristic principles and differential judgments in the assessment of information quality. *Journal of the Association for Information Systems, 18*(5), 403–432. http://doi.org/10.17705/1jais.00458

Batini, Carlo, Cappiello, Cinzia, Francalanci, Chiara, & Maurino, Andrea. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1–52. https://doi.org/10.1145/1541880.1541883

Berger, Charles, Blauth, Robert, Boger, David, Bolster, Christopher, Burchill, Gary, DuMochel, William, Pouliot, Fred, Richter, Reinhart, Rubinoff, Allan, Shen, Diane, Timko, Mike, & Walden, David. (1993). Special issue on Kano's methods for understanding customer-defined quality. *Center for Quality of Management Journal*, *2*(4).

Bowman, Alan. (2022). Getting personal: creating highly relevant content experiences with personalization. *Presentation at CIDM ConVEx 2022 Tempe.*

Carey, Michelle, McFadden Lanyi, Moira, Longo, Deirdre, Radzinski, Eric, Rouiller, Shannon, & Wilde, Elizabeth. (2014). *Developing quality technical information: A handbook for writers and editors* (3rd ed.). IBM Press.

Carroll, John, & van der Meij, Hans. (1996). Ten misconceptions about minimalism. *IEEE Transactions on Professional Communication*, *39*(2), 72–86. https://doi.org/10.1109/47.503271

Cheung, Iva. (2016). Four levels to accessible communications. *Personal blog*. https://ivacheung.com/2016/09/four-levels-to-accessible-communications/

Cichy, Corinna, & Rass, Stefan. (2019). An overview of data quality frameworks. *IEEE Access*, vol. 7, 24634–24648.

DeSanto, David (2025). Kano Survey for feature prioritization, *GitLab Handbook*, retrieved from https://handbook.gitlab.com/handbook/product/ux/ux-research/kano-model/

Eppler, Martin. (2006). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes* (2nd ed.). Springer.

Falla, Gerald. (2018). Personal communication.

Ge, Mouzhi, & Helfert, Markus. (2007). A review of information quality research—develop a research agenda. *MIT International Conference on Information Quality.*

Ge, Mouzhi, Helfert, Markus, & Jannach, Dietmar. (2011). Information quality assessment: validating measurement dimensions and processes. *ECIS 2011 Proceedings*, 75. http://aisel.aisnet.org/ecis2011/75

Gibbons, Sarah (2021). 5 prioritization methods in UX roadmapping, Nielson Norman Group. https://www.nngroup.com/articles/prioritization-methods/

Hauser, John. (1991). Comparison of importance measurement methodologies and their relationship to consumer satisfaction. *MIT Marketing Center Working Paper 91–1*.

Holst, Christian. (2012). UX and the Kano Model. Baymard Institute. https://baymard.com/blog/kano-model

Huang, Kuan-Tsae, Lee, Yang, & Wang, Richard. (1999). *Quality information and knowledge*. Prentice Hall.

International Organization for Standardization. (2015). Quality management systems — fundamentals and vocabulary (ISO Standard No. 9000:2015). International Organization for Standardization. https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en

Johnson, Tom. (2021). Different approaches for assessing information quality. I'd Rather Be Writing. https://idratherbewriting.com/learnapidoc/docapis_metrics_assessing_information_quality.html

Kahn, Beverly, Strong, Diane, & Wang, Richard. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM, 45*(4), 184–192. https://doi.org/10.1145/505248.506007

Kano, Noriaki, Seraku, Nobuhiko, Takahashi, Fumio, & Tsuji, Shin-ichi. (1984). Attractive quality and must-be quality. *Journal of The Japanese Society for Quality Control, 14*(2), 147–156. https://doi.org/10.20684/quality.14.2_147

Klein, Jeff. (2024). You're Wrong. My Documentation Is Brilliant! [Webinar], Society for Technical Communication Technical Editing Special Interest Group (STC TESIG).

Lee, Yang, Strong, Diane, Kahn, Beverly, & Wang, Richard. (2002). AIMQ: A methodology for information quality assessment. *Information & Management, 40*, 133–146. https://doi.org/10.1016/S0378-7206(02)00043-5

Lin, Feng-Han, Tsai, Sang-Bing, Lee, Yu-Cheng, Hsiao, Cheng-Fu, Zhou, Jie, Wang, Jiangtao, & Shang, Zhiwen. (2017) Empirical research on Kano's model and customer satisfaction. *PLoS ONE, 12*(9), e0183888. https://doi.org/10.1371/journal.pone.0183888

Löfgren, Martin & Witell, Lars. (2008). Two decades of using Kano's theory of attractive quality: a literature review. *Quality Management Journal*, *15*(1), 59–75. https://doi.org/10.1080/10686967.2008.11918056

Madzík, Peter. (2018). Increasing accuracy of the Kano model—a case study. *Total Quality Management*, *29*(4), 387–409. https://doi.org/10.1080/14783363.2016.1194197

Masycheff, Alex (2023). How measuring and managing content quality can help you prioritize your work. *Presentation at LavaCon 2023 San Diego*.
Mikulić, Josip (2007). The Kano model—a review of its application in marketing research from 1984–2006. *Proceedings of the 1st International Conference Marketing Theory Challenges in Transitional Societies*, 87–96.
Mikulić, Josip & Prebežac, Darko. (2011). A critical review of techniques for classifying quality attributes in the Kano model. *Managing Service Quality: An International Journal*, *21*(1), 46–66. https://doi.org/10.1108/09604521111100243
Moorman, Jan (2012). Leveraging the Kano Model for optimal results. *UX Magazine.* https://uxmag.com/articles/leveraging-the-kano-model-for-optimal-results
Mui, Allison & Shwer, Elle (2022). The art of surveys: analyzing data to improve the docs experience. *Product Gals blog*, retrieved from: https://medium.com/product-gals/the-art-of-surveys-analyzing-data-to-improve-the-docs-experience-6d8129ab9378
Olsen-Landis, Cary-Anne. (2017). Kano Model — ways to use it and NOT use it, IBM Design. https://medium.com/design-ibm/kano-model-ways-to-use-it-and-not-use-it-1d205a9cf808
OpenAI. (2023). ChatGPT-3.5 (24 May version) [Large language model]. https://chat.openai.com/share/dbd1c36b-0abd-424e-810b-3b2dc3caa2ed
Pipino, Leo, Lee, Yang, & Wang, Richard. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218. https://doi.org/10.1145/505248.506010
Richardson, Wendy. (2019). Get in the game—creating personalized customer experiences with a unified content strategy. *Presentation at LavaCon 2019 Portland.*
Snyder, Carolyn. (1998). Docs in the real world. User Interface Engineering. https://articles.uie.com/documentation/
Spool, Jared. (2013). Using the Kano Model to build delightful UX. *Presentation at Delight 2013 Portland*.
Spool, Jared. (2015). Is design metrically opposed? *Presentation at UXIM Salt Lake City*. https://www.uie.com/wp-assets/transcripts/is_design_metrically_opposed.html
Spool, Jared. (2018). Design metrics that matter. *Presentation at Business of Software Conference USA 2018 Boston*. https://businessofsoftware.org/2019/02/design-metrics-matter-jared-spool-uie-bos-usa-2018/
Stevens, Dawn, Madison, Kathy, & Ocker, Sabine. (2018). Metrics on metrics. *CIDM Best Practices*, *20*(6), 121, 125–132.
Strimling, Yoel. (2018). So you think you know what your readers want? *Intercom, 65*(6), 4–9.
Strimling, Yoel. (2019). Beyond accuracy: What documentation quality means to readers. *Technical Communication*, *66*(1), 7–29.
Strimling, Yoel. (2021). Editing for quality—from the readers' point of view. *CIDM Best Practices*, *24*(4), 1, 4–6.
Strimling, Yoel, & Michelle Corbin. (2009). Editing modular documentation: some best practices. *Intercom*, *56*(5), 29–31.
Strong, Diane, Lee, Yang, & Wang, Richard. (1997a). Data quality in context. *Communications of the ACM*, *40*(5), 103–110. https://doi.org/10.1145/253769.253804
Strong, Diane, Lee, Yang, & Wang, Richard. (1997b). 10 potholes in the road to information quality, *Computer*, *30*(8), 38–46. https://doi.org/10.1109/2.607057
Tontini, Gerson, Solberg Søilen, Klaus, & Silveira, Amélia. (2013). How do interactions of Kano model attributes affect customer satisfaction? An analysis based on

psychological foundations. *Total Quality Management & Business Excellence*, *24*(11–12), 1253–1271. https://doi.org/10.1080/14783363.2013.836790
Virtaluoto, Jenni, Suojanen, Tytti, & Isohella, Suvi. (2021). Heuristics revisited: developing a practical review tool. *Technical Communication*, *68*(1), 20–36.
Wang, Richard. (1998). A product perspective on Total Data Quality Management. *Communications of the ACM, 41*(2), 58–65. https://doi.org/10.1145/269012.269022
Wang, Richard & Strong, Diane. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099
Watts, Stephanie, Shankaranarayanan, Ganesan, & Even, Adir. (2009). Data quality assessment in context: a cognitive perspective. *Decision Support Systems 48*(1), 202–211. https://doi.org/10.1016/j.dss.2009.07.012
Witell, Lars, Löfgren, Martin, & Dahlgaard, Jens. (2013). Theory of attractive quality and the Kano methodology—the past, the present, and the future. *Total Quality Management and Business Excellence*, *24(*11–12), 1241–1252. https://doi.org/10.1080/14783363.2013.791117
Woodall, Philip, Oberhofer, Martin, & Borek, Alexander. (2014). A classification of data quality assessment and improvement methods. *International Journal of Information Quality*, *3*(4), 298–321. https://doi.org/10.1504/IJIQ.2014.068656
Zacarias, Daniel. (2015). The complete guide to the Kano Model. Folding Burrito. https://www.career.pm/briefings/kano-model
Zacarias, Daniel. (2018). Personal communication.
Yong, Louis. (2024). How do we communicate content quality? *Personal blog*. https://medium.com/@louis.yong.cc/how-do-we-communicate-content-quality-be01c0f91f65

Committee on Publication Ethics AI Guidelines Statement

As per the Committee on Publication Ethics (COPE) guidelines on AI and authorship (https://publicationethics.org/cope-position-statements/ai-author), the author states that the bulk of the material in the *Practical Applications* and *Conclusion and Future Directions* sections was written by the author; however, they include suggestions made by OpenAI's ChatGPT-3.5 LLM. The rest of this paper was written solely by the author.

Note from the TPC Series: The AI policy of the series is to not allow AI in any format. However, since this project was taken on from the STC, we followed their guidelines.