THE IMPORTANCE OF THE RAW IDEA IN INNOVATION: TESTING THE SOW'S EAR HYPOTHESIS

Laura J. Kornish
Leeds School of Business
University of Colorado
UCB 419
Boulder, CO 80309 USA
kornish@colorado.edu

Karl T. Ulrich
The Wharton School
University of Pennsylvania
500 Huntsman Hall
Philadelphia, PA 19104 USA
ulrich@wharton.upenn.edu

Original Version: November 2011 | This Version: October 2012

Abstract

This paper explores how important the quality of the raw idea is in determining success in innovation. On the one hand, one could argue that without a good idea, the chance of success is very small: "you can't make a silk purse out of a sow's ear." On the other hand, one could argue that with the right resources and approach, an innovator can create value out of just about anything: the Midas hypothesis. We provide a conceptual framework for thinking about this question generally and then test it empirically in one significant domain—household consumer products. We develop a novel data set from Quirky, a community-driven product development company. Our data include descriptions of the raw ideas originally proposed, the ultimate product designs that resulted from those ideas, and sales data. We augment the data from Quirky with multiple measures of idea quality that we obtained. We find that ideas do matter. We find that the quality of the raw idea as estimated by commercially feasible techniques is a significant predictor of market outcomes. In spite of the issues from selection and the variance introduced by measurement error, we do see a statistically significant relationship. We also conclude that for the domain we study, surveys of consumers are a better way to determine what a "good" idea is than ratings by even highly experienced experts. This is the first study we know of to demonstrate a significant link between idea quality and outcome using actual raw ideas and market outcomes.

Keywords: opportunity, opportunities, idea, raw idea, innovation, innovation process, product development, new product innovation, new products, product concept, evaluation, filter, selection, assessment, success, value, worth, performance

Acknowledgments: The Mack Center for Technological Innovation at Wharton and the Deming Center for Entrepreneurship at Leeds provided financial support for this research. Thomas Åstebro kindly shared data from his related work. We are grateful for helpful comments on a previous draft by Cheryl Druehl, Page Moreau, David Robertson, and participants at the 9th Annual Product and Service Innovation Conference.

1. Introduction

A "French fry restaurant" in a college town recently closed its doors. The short phrase "French fry restaurant" captures the idea underlying the eatery. It was a fast casual dining establishment that offered meals based on French fries: fries with sausage, fries with Thai toppings, fries with chili. Perhaps you are not surprised that this restaurant failed. You may be asking, "Is that restaurant concept even a good idea?"

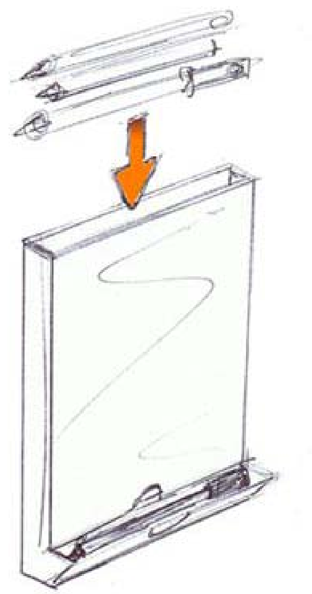
This paper explores how important the quality of the raw idea is in determining success in innovation. On the one hand one could argue that without a good idea, the chance of success is very small. The eighteenth-century author Jonathan Swift is reported to have said that "you can't make a silk purse out of a sow's ear." On the other hand one could argue that with the right resources and approach, an innovator can create value out of just about anything. (In fact, in 1921 engineers at consulting firm Arthur D. Little spun a silk purse out of processed sows' ears [Arthur D. Little, Inc., 1921].) This view might be called the *Midas hypothesis*, the extreme view of some venture capitalists that the team, the execution (and maybe luck) are everything (Ries 2011).

The *sow's ear hypothesis* is likely more valid in some domains than others. In pharmaceuticals, no matter what you do to the molecular compound, it won't become something it isn't intrinsically. Absent some revolution in alchemy, gold mines cannot extract gold from barren ore. However, perhaps an animated motion picture can be highly engaging even if the raw story line is mundane; or perhaps the story can evolve for the better as the movie is developed. In this paper, we provide a conceptual framework for thinking about how much raw ideas matter, and we believe that the framework is relevant to any domain. Then, we illustrate the framework by applying it in one significant domain—household consumer products.

We define an *idea* as an opportunity to create value through further investment (Terwiesch and Ulrich 2009). Ideas can take different forms in innovation. An idea may be the recognition of a new need—e.g., in the early 1990s, search engine innovators recognized the need to navigate the vast amount of information emerging on the worldwide web. An idea may be a new concept for a solution to an existing need—e.g., the idea underlying the *Snuggie* was a blanket with sleeves for the user's arms, a new way to stay warm. An idea may be the conjecture that an existing solution could meet an emerging need—e.g., the idea for the Macintosh computer was that the graphical user interface pioneered by Xerox as a corporate word-processing tool could address the burgeoning desire by individuals to access the power of computing.

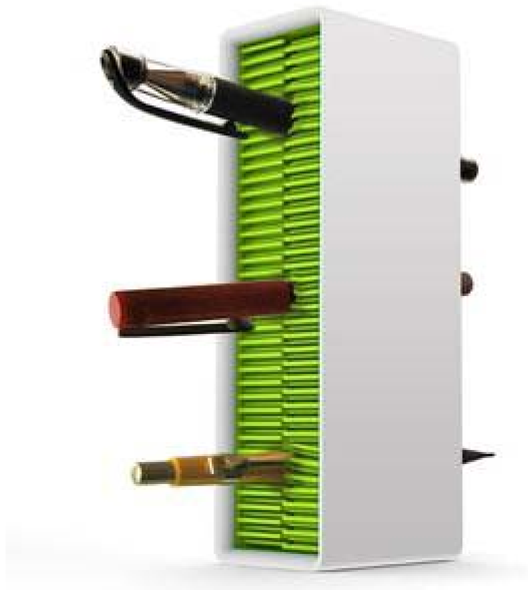
Ideas evolve over the course of the innovation process. We define the *raw idea* as the opportunity as conceived at the outset of the innovation effort in a specific organizational context. Of course in most cases, the raw idea as it first enters an organization's innovation process existed in an even more rudimentary state in the mind of the originator. A raw idea for an innovation is often expressed in words or with a simple visual depiction. For instance, Figure 1 shows an example of a raw idea from our data set, along with the corresponding final product that was developed from the raw idea. In our empirical setting, the raw ideas were the actual ideas proposed at the beginning of a structured innovation process. For other contexts, the raw idea would take a different form and the description might be more or less elaborated. For instance, in a pharmaceutical innovation process the raw idea might be a newly synthesized compound, described fully by its molecular structure. In a movie studio, the raw idea might be a one-sentence description of a plot.

Our conceptual framework recognizes that value from an innovation can originate in several sources: the idea itself, decisions that are made in the development and marketing of the idea, and exogenous factors. We examine the roles of these sources in creating value. With this framework, we can answer our central research question: how much does the raw idea matter in determining innovation success? We answer that by analyzing how much of the variation in outcomes in innovation is explained by variation in the quality of the raw ideas. Following on that question, we can also look at the best way to identify good ideas, given that idea quality is an elusive notion.



disPENser

We all have tons of pens lying around, mostly occupying often used drawers, or being stacked in some sort of funny cup-like holder... Either way it's a messy gang, most of them being cheap (usually promotional) biros, many of them dried up. To help maintain an overview of this Pen-o-mania, the disPENser is an outcome. It's no more than a thin (plastic or cardboard?) container, holding the pens in a neat orderly vertical fashion. The pens are stacked on top of each other, and when needed, you just grab the bottom one. When finished writing, put the pen back at the top, and this way all pens are used through time, instead of only the ones being nearby. The disPENser is dimensioned to the most common diameter of pens, so no, not all exotic pens will fit, but hey, that's why they're exotic in the first place. The front of the disPENser can get more functions, it could become a placeholder for the family-calendar, or just a plain notepad. I can see use in classrooms as well, where the disPENser would hold pencils or crayons (different dimensions apply off course). Hope you enjoy this idea, so let's start fighting these junk-drawers!



Pen Zen

Pen Zen is an elegantly designed storage unit for pens, pencils, highlighters, and other office supplies. Pen Zen's sleek form and functional supply holders make it easy to master the art of organization. Position your Pen Zen horizontally with the items sticking out of the top, or stand it upright with the items sticking out of the ends. It's your call; choose the arrangement that suits you best! Features:
- Rubber extrusions that act as magic fingers to hold your items in place.
- Glossy white plastic exterior makes for an appealing addition to your desktop decor.
With Pen Zen, your desk clutter can finally find inner peace. Dimensions: 7.5in x 2.5in x 3.0in.

Price: $30.00

Figure 1: Example raw idea (left) and final product design (right).

Empirically, the research question of how much the idea itself matters is challenging for several reasons. First, we face a measurement challenge. Idea quality is a theoretical notion, but we need actual measurements for our analysis. Second, we face a selection challenge. The vast majority of ideas never receive investment. Thus, we can only observe outcomes for a small fraction of ideas. This data limitation cannot readily be overcome experimentally because of the prohibitive costs associated with commercializing ideas and realizing their outcomes. Third, we face a data availability challenge. We need access to the raw ideas and to the outcomes (e.g., sales results) for a large sample of innovations. But, we need to be able to evaluate the quality of the ideas retrospectively, without polluting that evaluation with knowledge of the actual outcomes.[1]

In tackling the empirical challenges, we develop a novel data set. We use data from the community-driven product development company Quirky. Quirky runs weekly tournaments via its website, selects the best raw ideas, and leads a product development effort supported by the Quirky community. The successfully developed ideas are sold in an online store on the site. Due to the community orientation of the site, it is very transparent. The raw ideas are available on the site, and the sales figures for each product are updated as orders come in. We use multiple measures of idea quality, including the company's own ratings, purchase intent measures from a consumer panel, and expert ratings of the ideas.

Our findings are as follows. The answer to our central question is that ideas do matter. We find that the quality of the raw idea as estimated by commercially feasible techniques is a significant predictor of market outcomes. In spite of the issues from selection and the variance introduced by measurement error, we do see a statistically significant relationship. As a complement to answering our central question, we conclude that for the domain we study, surveys of consumers are a better way to determine what a "good" idea is than ratings by even highly experienced experts.

Why should academics and practitioners care about the role of the idea in innovation success? Organizations invest a lot of resources in opportunity identification and evaluation of ideas. If

[1] Chandy et al. (2006) describe this challenge of data availability, for both inputs and outputs, in their study of firms' abilities to convert ideas into launched products.

the answer to our research question is that ideas don't matter, or don't matter much, those resources could be better spent. If ideas don't matter, should innovators throw darts instead, or work with first idea generated, pivoting as needed? Or, if they matter a lot, then should they invest even more in the early stages of the product development process, identifying opportunities and carefully selecting them? This paper contributes to the academic literature on innovation by demonstrating that ideas do matter and providing evidence about valid ways to select them. The paper is organized as follows. First we discuss prior research in related areas. Then we lay out the conceptual framework; discuss the data, the analysis, and the results; and conclude.

2. Prior Work

To our knowledge this is the first work to examine the relationship between the quality of raw ideas as originally proposed and market outcomes. In studies of idea generation, there is an implicit assumption that better ideas have a significantly positive impact on better market outcomes, suggesting that firms should invest substantial resources in generating better ideas. No one posits that better ideas may lead to worse market outcomes— rather the question is really how important is the quality of the raw idea as a determinant of success. To date, there have been no published studies that empirically examine this question, using the both the original raw ideas and market outcomes. Our study fills that gap.

In particular, our work adds to the existing literature in three ways.

1. Unit of Analysis is Raw Idea as Originally Proposed

First, while other authors have studied the extent to which attributes and early evaluations of products predict market success (e.g., Rogers 1995, Goldenberg et al. 2001, Astebro 2003, Rogers 1995, Chandy et al. 2006, Kamakura et al. 2006, Eliashberg et al. 2007, Morwitz et al. 2007), none that we know of have used the raw ideas as originally proposed, and many are based on retrospective descriptions of the products. For example, in the work of Goldenberg et al. (2001), the studies are based on attributes coded from patents and retrospective synopses of ideas described in books. The work of Chandy et al. (2006), focusing on pharmaceuticals, also relies on patents as the basis for describing products. The work of Eliashberg et al. (2007) answers the question of whether movies with certain plot elements (i.e., a model of movie "quality") predict box office results. In that study, the “idea” is a 4-to-20-page narrative summary of the movie, created retrospectively. Similarly, in Study 2 of Morwitz et al. (2007), concept tests are shown to have predictive validity, however these “concepts” are descriptions of products ready to be launched. Other forms of market research, such as simulated test markets (e.g., Clancy et al. 2006), typically work with quite elaborated descriptions of products.

Although not based on actual ideas proposed in practice, one paper that relates raw ideas to market outcomes is Dahan et al.’s (2011) Securities Trading of Concepts (STOC) paper. The ideas are generated by the researchers and are described in simple visual and verbal depictions, based on levels of attributes. For one of the categories they study, cross-over vehicles, the authors examine the relationship between STOC evaluations and market shares. In that study, they do not find any statistically significant relationship between ideas and outcomes, a curious finding that motivates additional exploration of the research question.

Why is it important to understand the role of the quality of the raw ideas, as first proposed, rather than more fully developed ideas? Working with raw ideas represents a very real task that firms face, sorting through dozens, hundreds, or even thousands of possibilities typical of the “fuzzy front end” of innovation. Establishing the strength of the relationship between raw ideas and outcomes can help organizations make informed decisions about investments in idea generation and selection.

2. Dependent Variable is Market Outcomes

Second, while other studies have addressed various questions related to the quality of raw ideas, none that we know has used market outcomes as a dependent variable. Many studies related to idea generation track a dependent variable related to quality of the ideas, often with an evaluation by some expert panel. Goldenberg et al. (1999) used a panel of three senior marketers to evaluate ideas; Diehl and Stroebe (1987) used one or two research assistants to rate ideas on originality and feasibility; and Dahl and Moreau (2002) used panels of three experts to judge originality in one study and used 19 consumers in another, and also used panels of 4 or 16 consumers to indicate willingness to pay for raw ideas. Girotra et al. (2010) analyze the practices in the academic literature for judging quality of ideas. They conclude that the best ways to estimate idea quality are with holistic ratings of business value by trained experts and with purchase intent surveys of consumer panels. While all of these studies use the ideas as originally generated, none of them have an available market metric.

Why is it important to use market outcomes rather than other measures of quality? For both practitioners and scholars, the survey and judgment measures are a cheaper and more convenient proxy for market outcomes. However, the underlying assumption in using non-market measures is that they correspond to ultimate business value as would be determined in a market. We test this assumption. We acknowledge that we study only a particular empirical context, but our results provide some guidance about the relationship between widely used pre-market quality measures (namely, expert evaluations and consumer surveys) and market outcomes.

3. Relative Influence of Raw Idea and Final Product Design

Third, while other studies have looked at the relationship either between refined ideas and market outcomes (our first point above) or between raw ideas and quality judgments (our second point above), none that we know of look at how the stages of actual idea development contribute to value creation. Our novel data set allows us to examine not just the relationship between raw idea and market outcome, but to study two steps: first from raw idea to final product and then from final product to market sales. In so doing, we can conclude how much of the uncertainty about market outcomes is resolved with the design and development of the product based on the raw idea.

3. Framework for Innovation Success

By its very nature, innovation is a highly uncertain activity. Much has to happen to and around an idea before, ultimately, value is created, or not. We propose a general framework of value in innovation, capturing the essentials of product development. We call the framework the VIDE model (which we pronounce vee-day, Latin for “see”): V is value, I is the quality of the idea, D is how well the idea is developed and commercialized, and E is exogenous factors. These elements are conjunctive: if any one of them is wholly inadequate, little to no value will be created.

VIDE

To answer our central question, how much the quality of the raw idea matters in the success of an innovation, we examine the relationship between quality and success. Generically,

$$V = f(I, D, E) \qquad (1)$$

(We show the specific models we empirically estimate in the Analysis and Results Section.) In our study, we have measures of the idea quality (I). The exogenous factors (E) conceptually match an error term. To study the development issues (the D), we vary the context. First, we look at how well the quality of the raw idea predicts the quality of the final design, and second, we look at how well the quality of the final design predicts market outcomes. Figure 2 shows these two steps. By looking at the two steps, we can analyze how much the quality of the idea contributes at each step.

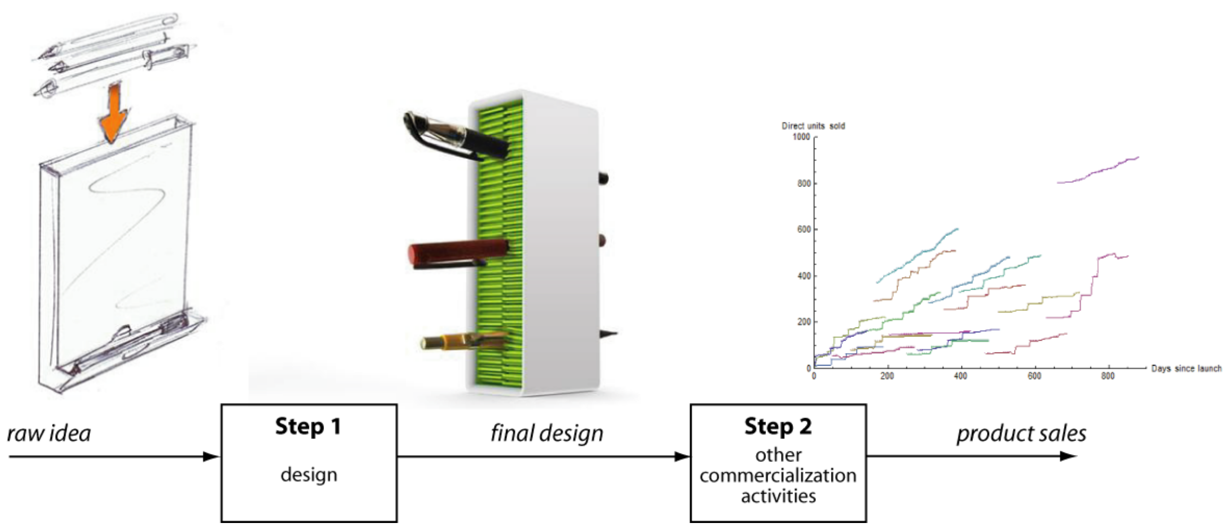


Figure 2: We analyze the relationship between the quality of the raw idea and the measure of success (sales rate) in two steps: from raw idea to final design and from final design to product sales.

Note that our *VIDE* models do not make any assumptions about the temporal sequence of activities in innovation. Indeed, *I*, *D*, and *E* could all be realized at the same point in time. However, in most innovation processes, there is a temporal sequence in which the factors are determined. The innovator starts with an idea, and therefore, an estimate of *V* based largely on knowledge about *I*. Then, the development process resolves further uncertainty. Finally, the commercialization process is completed and the exogenous factors realized to reveal an outcome.[2]

Defining the Quality of the Raw Idea

The true quality of a raw idea is a theoretical notion. Quality cannot be observed directly. All that can be observed is a particular realization of value for a particular set of outcomes of many random processes. We define the quality of the raw idea as a continuous variable reflecting the expected value to the innovating entity of pursuing that idea, given its particular context. Context is important: the quality of the idea of an "undo button in an elevator" will be higher for Otis Elevator than for IBM. In practice, while the true quality of an idea cannot be observed, it can be estimated. For instance, a panel of experts can rate the idea, or a survey of consumers can be used to measure purchase intent. For our work, we distinguish between the quality of the idea, I, and an estimate of that quality, $\hat{I}$. These two variables are related by $\hat{I} = I + \varepsilon$, where ε is the error in the estimate.[3]

Girotra et al. (2010) examine multiple dimensions of idea quality (technical feasibility, novelty, specificity, demand, and overall value). They conclude that multiple dimensions of quality load on a single factor and are highly correlated with holistic assessments and purchase intent measures, establishing that a single readily estimated metric can capture multiple dimensions.

Rust et al. (2002) and Golder et al. (forthcoming) discuss the concept of quality more generally, beyond just idea quality. Our notion of quality relates to the revenue-enhancing activity in Rust et al. (2002), as it springs from product innovation. For the same reason, it spans the quality production and perception processes described in Golder et al. (forthcoming).

---

[2] Of course we have not considered all nuances in estimating value. For instance, there can be complications associated with pricing, timing of cash flows, and kill options.

[3] If the error is multiplicative, $\hat{I} = I\varepsilon$, then the additive formulation holds when the equation is logged.

Variation Explained by *I*

Our central question, posed in terms of our model, is how much of the variation in V is explained by variation in I? This is a question that can be answered with the R^2 statistic from a linear regression. Note that R^2 is the squared correlation of V and I. We also examine the partial R^2, the fraction of variance explained when the effect of controls is removed.

The variability of I compared to the variability of the other factors is a measure of the signal value in the idea distribution. (Here we use "signal value" in the sense of signal-to-noise ratio.) In the extreme, if every idea were the exact same quality (i.e., no signal value), the quality of the idea wouldn't matter at all.

One of the empirical challenges to our research question is that only a small fraction of ideas is ever developed, and the fraction is selected in a systematic way; only the ideas deemed best are developed. This censoring has two effects. First, it reduces the sample size of ideas compared to a situation in which every idea is developed. Second, it changes the variance in idea quality in the observed sample. Sackett and Yang (2000) explain how this restriction in range, either through strict truncation of the independent variable or through a correlated variable used for selection, typically reduces the variance in the sample, the correlation, and therefore the measure of variance explained. We investigate the role of these effects in our data.

4. Data

Our data set comprises raw ideas and final products from a product development and commercialization company, independent evaluations of those ideas and products, and market outcome measures for the products.

Company

Quirky.com is a community product development website. The company specializes in "consumer products that could retail for under $150 and don't involve integrated software," including products both for the home (e.g., kitchen accessories) and the office (e.g., products to keep electronic devices organized). Quirky runs weekly contests in which community members contribute ideas. The ideas are described with text and/or images, and the best idea or ideas are selected from each contest. Moving through the product development process, community members contribute to market research, product design, and naming. Members earn points for participating in the process, and then earn money based on those points and product sales. The development is not completely crowdsourced; Quirky employees are heavily involved in the final selection of ideas and the actual development and production arrangements for the products.

The site was attractive as a data source for our research questions because all of the ideas from the weekly contests were publicly displayed on the site. In addition, the sales figures for the products in the store were clearly displayed. The transparency springs from the community involvement in the site: community members have an interest in tracking the sales progress of products they have contributed to.

Alexa.com reports the demographics for the quirky site: "Based on internet averages, quirky.com is visited more frequently by users who are in the age range 25-34, have no children, are college educated and browse this site from work" (Alexa.com, 2011).

Ideas

Our data set comprises the 101 products that were available in the store as of June 2011. We retrieved ("scraped") from Quirky's site the description of the product in the store and images of the final products. We have sales figures starting in March 2011 for the products that were in the store then; other products were introduced as we watched. We collected sales figures on a regular basis between March and November 2011.

We also retrieved the "raw idea" (text and images, if available) associated with each product in the store. There were 98 such raw ideas; a few of the 101 products were developed from the same raw idea. Part way through our observation period, the site began posting the Quirky staff's rating assigned to the finalist ideas in the weekly contests. We have this Quirky Score for 39 of the 98 raw ideas. (The Quirky Score is actually a composite of five sub-scores named *community*, *staff, design, market,* and *viability*.) Figure 1 contains an example of a raw idea and a final product design in our data set.

Finally, we also generated a random sample of 100 raw ideas from all the weekly contests, that is, the entire population of ideas on the site, not just ideas that were developed. (Exactly one of these 100 ideas was, in fact, selected for the store, consistent with the approximately one-percent selection ratio in the Quirky contests overall.)

Success Variables

We tracked units sold in the store and recorded prices for each of the products, allowing us to look at both sales volume and revenue for the products. Quirky breaks out the unit sales into direct and indirect units; direct units are sold through the Quirky website; indirect units are sold on a wholesale basis to retail partners. The prices for all the products were essentially constant over the observation period. A few of the products had a few observations with different prices. We used the median prices in the observations, which, in every case, is also the modal value.

Products were introduced to the store continually, and therefore at different times. Quirky reports not only units sold but also days in the store. The introduction timing raises the question of whether we can compare unit sales (or revenue) for a product that has been in the store 30 days with one that has been in the store 330 days. To normalize for the length of time in the store, we look at sales *rates*. Inspection of the sales trajectories reveals that they are reasonably linear. Figure 3 shows a sample of sales trajectories in the data.

Because we do not know manufacturing costs, we do not know the exact profit margins of each product. However, the gross margins for a direct-to-consumer specialty retailer of its own proprietary household products are typically quite high (i.e., >75%) so revenues are very highly correlated with profits (Ulrich and Eppinger, 2011). In our main analysis, when we use sales rate or revenue as a measure of success, we are implicitly assuming that the margin percentage is the same across all the products. In our analysis, we check the robustness of this assumption with estimates of manufacturing costs.

In sum, we track the following success variables or outcomes: direct units sold, total units sold, sales rate, revenue, and revenue rate. These measures are consistent with the results of the Product Development Management Association (PDMA) Taskforce on Measures of Success and Failure (Griffin and Page, 1993 and 1996).

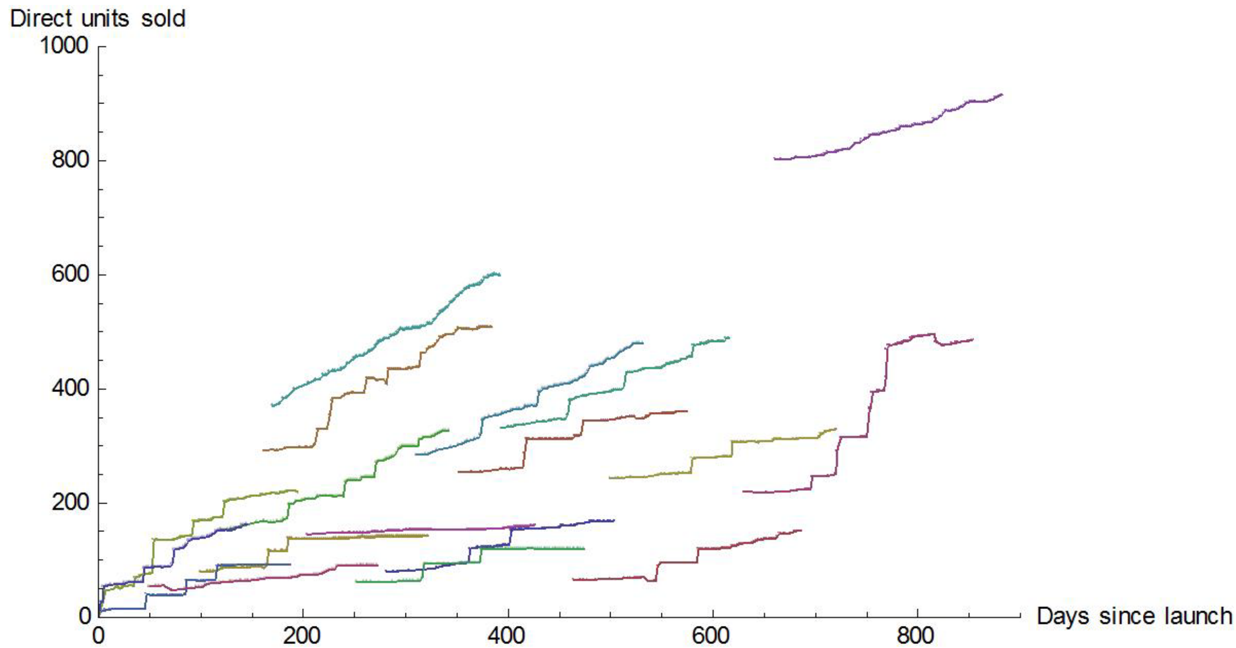


Figure 3: Sales trajectories for 20 of the products in the sample. Drops in units sold reflect returns. The trajectories shown are the sales over the observation window from March 25, 2011 to November 17, 2011.

Measures of Idea Quality

Another key variable for our research question is the quality of the raw idea. We obtained four measures of quality.

The first measure of idea quality is the Quirky Score itself. As explained above, this is Quirky's internal scoring scale for evaluating ideas. Although we only have the Quirky Score for some (39) of the ideas, the missing data do not create a selection bias; the reporting change was unrelated to idea quality. Rather, it simply represented a change in communication policy.

The second measure of idea quality is purchase intent for the raw idea. Probability of purchase conditional on awareness and availability, for a given price, is essentially a measure of the quality of the idea. The better the solution relative to the alternatives and the more pervasive the need it addresses, the more likely a user in the target market is to purchase the innovation. Purchase intent was one of the recommended measures from the study of idea quality by Girotra et al. (2010). Moore (1982) documents that concept screening using a purchase intent question is an established practice in industry.

We measured purchase intent for the 98 raw ideas that made it into the store and also for the 100 random raw ideas drawn from the entire population of ideas on the site. Paid online panelists looked at product descriptions (text, and if provided by the originator of the idea, an image) and rated their purchase intention on the typical five-point scale (definitely not, probably not, might or might not, probably, definitely). We used an attention filter question to screen out people who were not reading the survey. Each panelist rated 50 ideas, and each idea was rated an average of 107 times. The purchase intent responses were translated to a single overall purchase-intent score by weighting each of the responses with 0, 0.25, 0.5, 0.75, 1, respectively (Jamieson and Bass, 1989). Consistent with typical practice for testing raw ideas, for which design concepts are not yet developed, the product descriptions did not suggest prices.[4]

The third measure of idea quality is purchase intent for the final design. We followed a similar procedure to obtain these measures. A different group of paid online panelists looked at product descriptions and images of the final products, as portrayed in the (online) Quirky store. They answered a priced purchase-intent question using the same five-point scale as above. As above, each panelist rated 50 ideas and faced an attention filter question. Each idea was rated an average of 89 times. The descriptions of the final products did include prices, as prices were set as part of the final development process.

The fourth measure of idea quality was ratings by experts. We used seven experts in consumer-products marketing and product development to rate each of the 98 store ideas plus fifty out of the set of 100 random ideas. (All seven experts saw the same fifty.) The experts each had at least 15 years of experience in designing, developing, or commercializing consumer products. The experts all have experience in multiple product categories over the course of their career, with great overlap in the set of categories in our data set. The qualifications of our experts compare favorably to those reported in the literature (e.g., Goldenberg et al. 1999, Diehl and Stroebe 1987, Dahl and Moreau 2002, Girotra et al. 2010).

---

[4] Ottum (2005) explains this common practice, “It is usually a good idea not to put a price on …early ideas, because the goal of concept testing is to get a read on customer interest in the general idea” (p. 295).

Experts were asked to rate the ideas on a scale of 0 to 10, based on anticipated units sold. The question to expert was phrased in terms of units sold to make it comparable to the consumer survey. (The purchase intent measure yields an estimate in terms of units.) As with the raw ideas shown to consumers, prices are unavailable, and we acknowledged that lack of availability directly with the instruction that the experts assume the “resulting products would be priced appropriately.”

One of the challenges in finding a suitable data set for this research question is that raters of idea quality need to be unaware of the actual resulting products and their commercial success. Our analysis relies on the fact that the Quirky market is limited and relatively unknown. We could not take the same measures for products sold in mass, mainstream channels like Target and Wal-Mart. That availability would pollute our ability to go back and measure idea quality. In our purchase-intent surveys, we verified that our respondents were not biased by knowledge of market outcomes by asking their familiarity with a set of on-line retailers which included Quirky.com. Very few (less than 2%) had heard of Quirky; we eliminated those responses.

In the measures described above, we have addressed *V* (success measures), *I* (idea quality measures), and *D* (purchase intent for raw ideas vs. final products) in the *VIDE* framework. Another attractive feature of using the Quirky data is that the single development and sales platform controls, to some extent, the variation in exogenous factors, *E*. There are still exogenous factors that affect outcomes for the products in our sample, but some of those, such as the size of the aware population, are relatively constant for all the products on the site.

5. Analysis and Results

In this section, we present the answers for our research questions. Does the quality of the idea matter? If so, how much? And what’s the best way to measure that quality?

Purchase Intent and Sales Rate

We answer our central question about the importance of the raw idea by estimating the relationship between purchase intent (PI) for the raw idea (a measure of raw idea quality) and sales rate (a measure of success). Our data allow us to not only test that question directly, but to decompose the analysis into two steps: first, the relationship between purchase intent for the raw idea and purchase intent for the final product, and second, the relationship between purchase intent for the final product and sales rate. See Figure 2. In our main analysis, we use ordinary least squares (OLS) to estimate Equation (2) and Equation (3) below, in which p represents the control variable, price.

$$\mathrm{Ln(Sales\ Rate)} = \beta_0 + \beta_1 I + \beta_2 \mathrm{Ln}(p) + e \qquad (2)$$

$$\mathrm{PI\text{-}Final\ Design} = \beta_0 + \beta_1 I + \beta_2 \mathrm{Ln}(p) + e \qquad (3)$$

By adopting a different measure for I, Equation (2) can test both Step 2 (from Figure 2, i.e., final design to outcome) as well as both steps together (raw idea to outcome). Equation (3) can test Step 1. We expect a negative relationship between the price of the product and the quantity sold.

One natural question to ask about including price as a control is whether price is endogenous, and therefore raises econometric issues. We present our main analyses using OLS with price as a control variable, but we also address the endogeneity issue. To check whether price is correlated with the regression error (violating an OLS assumption), we use estimates of manufacturing costs as an instrument for price and apply a Hausman test. We report this additional cost analysis at the end of this section.

Another natural question is whether we need to control for the product category of the idea. For example, do pet products need to be treated differently than kitchen products? In this data set, our judgment is that such a control is *not* appropriate. Quirky develops products across categories. They are looking for the best ideas regardless of category, within their scope of consumer products. Therefore, the relevant question for them is the one we analyze: what's the relationship between the quality of any idea and success? Category is relevant to the extent that it influences cost and price, and we do account for those factors.

Model 1 in Table 1 shows the results of regressing log sales rate on purchase intent of the raw idea. This is the most basic test of the relationship between idea quality and success. We observe that the raw idea is a statistically significant predictor of sales rate, and the variation in raw idea quality as measured by purchase intent explains 4.3 percent of the variation in log sales rate. (Throughout our analyses, we use the natural logarithm of the sales rate and the price: these two variables have values that extend over more than a factor of ten, with a long right tail. Table 1 shows analyses with the non-logged versions of the idea quality measures. The results are not substantially different when those quantities are logged too.)

We include a version of each model that controls for price. In Model 2, we observe that the quality of the raw idea, as measured by purchase intent, remains statistically significant as a predictor of success when controlling for price. The partial R^2 on quality in Model 2 is 0.07. (For space considerations, we did not include all the partial R^2s.) We include Model 3 as a baseline; log price alone explains 16 percent of the variation in log sales rates—inexpensive products simply sell at higher rates.

In Model A, we examine Step 1 from Figure 2. We test whether the purchase intent of the raw idea predicts purchase intent of the final product design. Estimating purchase intent of final product design based on purchase intent of the raw idea yields an R^2 of 0.20, and controlling for price in Model B, the R^2 is 0.54, with a partial R^2 on quality in Model B of 0.35.

In Models 4 and 5 we examine Step 2 from Figure 2. We test the relationship between purchase intent of the final design and sales rate. Without controlling for price (Model 4), the R^2 is 0.17 and when controlling for price (Model 5), the R^2 is 0.21. Note that the additional explanatory power of price is modest in this case, most likely because the purchase intent measure using the final product design included the retail price in the description of the product.

In Models 1, 2, A, and B, the purchase intent independent variable is positive and significant. These results give us the answer to our central question: yes, the quality of the idea, as measured by purchase intent, is a significant predictor of better outcomes. However, the comparison of the two steps (Model A, with PI-raw to PI-final, and Model 4 with PI-final to sales) is also telling. The R^2 of the first step (0.20) is comparable to the second step (0.17). The estimated quality of the raw idea explains 20 percent of the variation in how much consumers will like the final product design, with price accounting for a large fraction of the remaining variance. However, even after the product design is fully determined, its quality as measured by purchase intent explains only 17 percent of the variation in eventual success. Thus, a large amount of variation in outcomes is determined by factors other than the idea and the quality of the product design based on that idea. The net result is that the estimated quality of the raw idea itself only explains 4.3 percent of the variation in success.

Using our regression results, we can calculate the effect on sales rate of starting with a better raw idea. The standard deviation of purchase intent (based on the raw idea) in the *random* sample of ideas is 0.0843. Using Model 2 from Table 1, a one standard deviation change in purchase intent translates to a 0.255 change in log sales rate (0.255 = 0.0843*3.020). In other words, the sales rate goes up 29% ($e^{0.255}$ -1) with a one-standard-deviation increase in purchase intent.

Table 1: Results of regression analyses. Standard errors in parentheses. (*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Log Sales Rate

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Constant	-1.154**	0.721	1.764***	-1.567***	0.132	-0.212	1.815***	0.836	2.860***
	(0.470)	(0.582)	(0.455)	(0.323)	(0.774)	(0.374)	(0.579)	(0.723)	(0.771)
Log(Price)		-0.623***	-0.591***		-0.375**		-0.592***		-0.683***
		(0.132)	(0.135)		(0.156)		(0.136)		(0.163)
PI-Raw Idea	2.562**	3.020***							
	(1.209)	(1.101)							
PI-Final Design				4.455***	2.963**				
				(0.995)	(1.154)				
Ave. Expert Rating						0.007	-0.011		
						(0.084)	(0.078)		
Quirky Score								-0.013	-0.009
								(0.013)	(0.011)
N	101	101	101	101	101	101	101	39	39
R^2	0.043	0.22	0.16	0.17	0.21	0.00	0.16	0.03	0.35
Adj. R^2	0.034	0.21	0.15	0.16	0.20	-0.01	0.14	0.00	0.31

Dependent Variable: Purchase Intent of Final Design

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Constant	0.117***	0.354***	0.551***	0.252***	0.498***	0.358***	0.579***
	(0.039)	(0.041)	(0.039)	(0.034)	(0.048)	(0.076)	(0.079)
Log(Price)		-0.079***	-0.073***		-0.072***		-0.074***
		(0.009)	(0.011)		(0.011)		(0.017)
PI-Raw Idea	0.512***	0.570***					
	(0.102)	(0.078)					
Ave. Expert Rating				0.014*	0.012*		
				(0.008)	(0.006)		
Quirky Score						-0.000	-0.000
						(0.001)	(0.001)
N	101	101	101	101	101	39	39
R^2	0.20	0.54	0.29	0.03	0.31	0.00	0.36
Adj. R^2	0.20	0.53	0.28	0.02	0.30	-0.02	0.32

Expert Ratings

Next we examine estimates of idea quality based on the ratings from our seven experts. The experts exhibited a fairly low level of agreement with one another. The correlation matrix for the seven experts and their average is shown in Table 2. A commonly used measure of interjudge reliability, Cronbach's alpha, is 0.55, which is considered poor agreement (Cronbach 1951).

Table 2: Correlation matrix of expert ratings on 148 raw ideas plus correlations with log sales rate for the 101 ideas that were launched.

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Average
Expert 1	-							
Expert 2	0.273	-						
Expert 3	0.082	0.117	-					
Expert 4	0.316	0.132	0.020	-				
Expert 5	0.134	0.243	0.055	0.119	-			
Expert 6	0.088	0.204	0.045	0.226	0.159	-		
Expert 7	0.181	0.136	0.124	0.010	-0.026	0.030	-	
Average	0.532	0.556	0.335	0.514	0.599	0.492	0.431	-
Log Sales Rate	-0.043	0.023	-0.022	-0.044	0.161	-0.027	-0.098	0.008

Is the average of the experts more informative than any one expert? Only Experts 2 and 5 outperform the average of the seven experts, as measured by correlation. The last row of Table 2 contains the correlations between the expert ratings and log sales rate.

The experts' ratings of the raw ideas that were developed and sold in the store are *not* a statistically significant predictor of sales rate, even controlling for price. These regression results are in Models 6 and 7 in Table 1. Even Expert 5's ratings by themselves are not a statistically significant predictor of sales rate (regression results omitted from the table). The average expert rating does predict purchase intent of the final design in a marginally significant way (Models D and E), with an R^2 of 0.03 for Model D.

The low level of expert agreement is consistent with the null result; from psychometrics we know that measures won't have predictive validity if they don't have reliability. With our consumer-based measures, agreement is not an issue. We are asking the consumers their personal likelihood of purchase; there is no reason to expect agreement. Further, because we have a much bigger sample of consumers (compared to the experts), we get a better measure of the overall market attractiveness of the ideas.

Quirky Scores

The Quirky Scores are the scores that the Quirky site uses for its own selection process. The site only started posting the scores on the site after a certain date, so the sample size is smaller (N=39) for this analysis than for those using the other idea quality measures. We do not see a statistically significant relationship between sales rate and Quirky Score, even controlling for price. These results are in Models 8 and 9 in Table 1. Similarly, we do not find statistically significant predictive power for the Quirky score on purchase intent of the final design (Models F and G).

In our study, the two populations of ideas—the ones that made it into the store and the ones that were randomly selected from the entire pool of raw ideas—have statistically significantly different values for purchase intent ($\mu_{store} = 0.38$ and $\mu_{random} = 0.34$, $p < 0.01$) and for the experts' ratings ($\mu_{store} = 4.26$ and $\mu_{random} = 3.37$, $p < 0.01$). Thus, the Quirky filter is meaningful, even if the Quirky score itself is not a statistically significant predictor of outcomes in our data.

This lack of statistical significance in Models 8, 9, F, and G could easily be due to the smaller sample size. Another plausible explanation for the lack of statistical significance is the restriction in range of the independent variable. The 39 ideas in the sample are among the very best of the Quirky score distribution,[5] and may therefore provide little range, and more importantly, little variance, with which to make statistical inferences in the regression analysis. If we could estimate the variance in Quirky scores for the whole population of ideas, we could correct our estimates for range restriction, and may possibly see significance.

[5] Ideas are selected in weekly contests, so the selected ideas do not strictly form the upper tail of the distribution.

Quirky scores are only reported for the selected ideas, not all the ideas, so we need some way of estimating the Quirky score distribution for the whole population of ideas. To estimate that distribution, we consider five plausible distribution shapes: Normal, Lognormal, and the three extreme value distribution forms, namely Weibull, Frechet, and Gumbel. (Dahan and Mendelson, 2001, explain that extreme value distributions are useful in modeling the value from innovations.). We estimate the two parameters for each of these distributions by simulating the Quirky selection process of choosing the best idea from weekly contests, and finding the parameter values that yield matches to the observed selected Quirky score mean (56.4) and standard deviation (10.5).

The best matches for the Lognormal, Weibull, and Frechet distributions yield standard deviations of 9.1, 11.5, and 5.7. The best matches for Normal and the Gumbel distributions yield negative mean scores, so we eliminate them from further consideration. The interesting thing to note is that two of these possible values (9.1 from Lognormal and 5.7 from Frechet) are lower than the observed standard deviation in the selected ideas (10.5). The third value (11.5 from Weibull) is not much greater.[6]

The threat to statistical power in a systematically selected sample comes from reduced variance in the sample. However, the evidence we see suggests that the variance in the sample is either not restricted at all, or barely restricted, compared to the population. Therefore, the failure of the Quirky scores in predicting outcomes cannot be explained by reduced variance, and the established corrections for attenuation due to restriction in range (see Sackett and Yang 2000 for the typology) do not apply.

Other Measures of Success

In the analyses above, we use sales rate as the dependent measure. We have several other possible measures of value, including units sold, revenue, and revenue rate. Our results are generally robust to different versions of the analyses.

---

[6] Other measures of quality show a similar pattern. For purchase intent for the raw idea, the standard deviation of the selected sample and the random sample are both 0.084. For the experts, the standard deviations are 1.22 and 1.21, for selected and random samples, respectively. (Note that the scale for purchase intent is 0-1, and for experts 0-10.)

The correlation matrix Table 3 shows the relatively strong positive correlation among the different outcome measures (units, total units, sales rate, revenue, and revenue rate). It also shows the relatively weak (and in some cases negative) correlations among the measures of quality themselves and between the quality measures and the outcomes. The purchase intent measures tend to have higher correlations with outcomes than the expert evaluations, both those of our experts and of the Quirky staff. Finally, we note that the strength of the correlation of the purchase intent measure based on the final product with the outcome measures. This pattern confirms that significant uncertainty about the idea has been resolved when the design is finalized and the price set. In fact, this purchase intent measure can even be thought of both as a measure of idea quality after it has been developed, and, in some sense, as an outcome measure for the development process.

Table 3: Correlation coefficients among price, success measures, and estimates of idea quality. Price and success measures are logged; measures of quality are not. Correlations with Quirky score contain only 39 observations, compared to 101 for the others. With correlation above 0.195 and 101 observations, the relationship is significant at the 0.05 level. (For 39 observations, the correlation has to be above 0.315.)

	Price	Units	Total Units	Sales Rate	Revenue	Revenue Rate	PI Raw Idea	PI Final Design	Ave. Expert Rating	Quirky Score
Price	-									
Units	-0.47	-								
Total Units	-0.38	0.76	-							
Sales Rate	-0.40	0.83	0.62	-						
Revenue	0.28	0.72	0.53	0.59	-					
Revenue Rate	0.29	0.53	0.37	0.76	0.81	-				
PI Raw Idea	0.09	0.05	0.17	0.21	0.12	0.28	-			
PI Final Design	-0.54	0.33	0.36	0.41	-0.07	0.05	0.45	-		
Ave. Expert Rating	-0.05	-0.02	0.04	0.01	-0.06	-0.03	0.34	0.18	-	
Quirky Score	0.10	0.03	-0.01	-0.17	0.15	-0.10	0.27	-0.06	0.19	-

Cost Estimation

As we explain above, price can be considered endogenous. The endogeneity arises because cost may influence both price and sales rate. We use an instrumental variables approach to address possible endogeneity. For the final products, the images and descriptions reveal information about the materials, size, number of parts, and types of parts (e.g., whether the product contains electronics), major drivers of cost. We estimate of costs based on these factors, as prescribed by Ulrich and Eppinger (2011).

Cost estimates serve as an instrumental variable for price. The correlation between log(cost) and log(price) is 0.78. Following the standard instrumental variable procedure, we first estimate price as a function of quality and cost, and then use the estimated price in the original regression, either sales rate for Models 2, 5, 7, and 9 or purchase intent for the final design for Models B, E, and G. We also estimate Models 3 and C, which omit quality, serving as a baseline to isolate the effect of price.

Table 4 shows the results of this two-stage least squares (2SLS) analysis. The Hausman test can be used retrospectively to determine whether endogeneity is truly a problem. We run the version of the test that augments the original regressions (i.e., the ones from Table 1) by including the residuals from the first stage (log price on quality and log cost) as an independent variable. If the coefficient on the residuals is significant, then endogeneity is an important issue to consider, otherwise, not (Wooldridge 1999). Our results show that for the models that do not use purchase intent on the final design, Models 2, 3, 7 and 9, the coefficient on the first stage residuals is not significant, meaning that the Hausman test shows that endogeneity of price is not a concern. For the models that do include purchase intent on the final design, endogeneity is an issue, and the 2SLS analysis is preferred.

Comparing the results of Stage 2 in Table 4 to the results in Table 1, we see that the signs of the coefficients and the level of significance are nearly all the same and the R^2s are all very close. The exception is Model 5. In the 2SLS version, the purchase intent on the final design is no longer a significant predictor of sales (see Stage 2). Stage 1 of that models shows that quality *is* a significant predictor of price. Therefore, the quality information is embedded in the price information. We note that the descriptions of the final designs did include price.

Table 4: Results of two stage least squares regression analyses. Standard errors in parentheses. (*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Log Sales Rate

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Stage 1: Regress log price on log cost, and for most models, quality measure (PI-Raw, Ave. Experts, or Quirky Score)									
Constant		2.522***	2.494***		3.413***		2.784***		2.365***
		(0.211)	(0.078)		(0.157)		(0.166)		(0.401)
Quality		-0.077			-2.643***		-0.071*		-0.000
		(0.537)			(0.409)		(0.036)		(0.007)
Cost		0.529***	0.528***		0.463***		0.535***		0.612***
		(0.044)	(0.043)		(0.038)		(0.043)		(0.085)
Stage 2: 2SLS estimation using estimated log price									
Constant		1.147*	2.135***		1.196		2.188***		2.675***
		(0.670)	(0.583)		(0.972)		(0.688)		(0.871)
Estimated		-0.764***	-0.704***		-0.610***		-0.701***		-0.620***
Log(Price)		(0.171)	(0.175)		(0.203)		(0.174)		(0.212)
PI-Raw Idea		3.123***							
		(1.110)							
PI-Final Design					2.028				
					(1.272)				
Ave. Expert Rating							-0.015		
							(0.078)		
Quirky Score									-0.009
									(0.011)
R^2		0.22	0.16		0.20		0.16		0.35
Hausman Test: OLS of dependent variable on quality measure, log price, and first stage residuals									
First Stage		0.353	0.284		0.599*		0.284		-0.153
Residuals		(0.268)	(0.276)		(0.316)		(0.280)		(0.336)
N		101	101		101		101		39

Dependent Variable: Purchase Intent of Final Design

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Stage 1: Regress log price on log cost, and for most models, quality measure (PI-Raw, Ave. Experts, or Quirky Score)							
Constant		2.522***	2.494***		2.784***		2.365***
		(0.211)	(0.078)		(0.166)		(0.401)
Quality		-0.077			-0.071*		-0.000
		(0.537)			(0.036)		(0.007)
Cost		0.529***	0.528***		0.535***		0.612***
		(0.044)	(0.043)		(0.043)		(0.085)
Stage 2: 2SLS estimation using estimated log price							
Constant		0.288***	0.463***		0.419***		0.543***
		(0.048)	(0.051)		(0.058)		(0.090)
Estimated Log(Price)		-0.057***	-0.046***		-0.049***		-0.062***
		(0.012)	(0.015)		(0.015)		(0.022)
PI-Raw Idea		0.554***					
		(0.080)					
Ave. Expert Rating					0.012*		
					(0.007)		
Quirky Score							-0.000
							(0.001)
R^2		0.53	0.29		0.31		0.36
Hausman Test: OLS of dependent variable on quality measure, log price, and first stage residuals							
First Stage Residuals		-0.055***	-0.067***		-0.060***		-0.030
		(0.018)	(0.023)		(0.023)		(0.034)
N		101	101		101		39

6. Discussion

In addition to answering our central research question, essentially a scientific question about how much the idea matters, our results have practical implications for product development processes, particularly the "fuzzy front end" (Hauser et al., 2006).

Managerial Implications

The first implication is that ideas matter. Conventional wisdom among some entrepreneurs is that the idea doesn't matter: the team matters and execution is everything. At the root of this "wisdom" are two truths. First is the understanding that judging the quality of an idea is difficult. Second is that the conjunctive nature of success limits the influence of the idea; the idea *and* the design *and* the marketing *and* the market conditions all matter. We agree that measurements of idea quality are noisy and that idea quality is only part of the picture, but our results show that idea quality does, in fact, predict outcomes. Just because it's hard to know which ideas will be successful doesn't mean the idea itself doesn't matter.

The second implication is about the value of idea selection. Because having a good idea leads to a greater level of success, there is value in accurate selection. In the work of Dahan and Mendelson (2001), which derives the optimal number of concepts to test, early-stage evaluations are assumed to be unbiased, even if noisy, estimates of value. Even if the early-stage evaluations are unbiased estimates of true quality, there is a loss in value associated with the noise due to investing in ideas that appear to be the best ones, but are actually not.

The third implication is about how to measure idea quality. In our analysis, using a crowd of ordinary consumers to state purchase intent was a better gauge of market outcomes than a panel of experts. This finding builds on work such as Hoch (1988), which shows that experts have insufficient knowledge about the activities, interests, and opinions of American consumers; Tetlock (2005), a long-term study of the predictive power of experts; and other studies like those of Faulkner and Corkindale (2009); and Ozer (2009). Our experts are product designers, marketers, and consultants with years, and in some cases decades, of experience. The low level of agreement among experts (cf. Kamakura et al. 2006) is a clue that individual experts or even a panel of experts are insufficiently prescient. We find that there is wisdom in a crowd of consumers.

Caveats and Limitations

Several caveats and limitations bound the implications of our results.

The domain of consumer housewares is a multi-billion dollar industry, and so is economically significant in its own right. However, we would not expect the specific numerical results from this domain to apply in, say, movies or pharmaceuticals. This is a fundamental limitation of most empirical research. For example, the work of Chandy et al. (2006) is an empirical study of the pharmaceutical industry, but offers conceptual insights about conversion ability more generally. On the one hand the only way to control enough variance in the setting to get reasonable results is to focus on one domain or even one firm. On the other hand, that focus carries with it idiosyncratic factors that limit the ability to generalize the numerical results. We hope that the model of success and the associated insights are useful generally, and that the consumer housewares setting provides both a concrete illustration of the model and some numerical results that are useful to those interested in consumer goods.

In our Introduction, we acknowledge that the effects we observe in our data may very well differ by industry. For example, pharmaceuticals may be more sow's ear than Midas, with value inherent to and immutable in the idea.

The Quirky data are idiosyncratic for other reasons as well. The open nature of the Quirky development process is what allows us to study it. However, most commercial innovation processes are closed to outside observers. Some of Quirky's customers are also participants in its development process. One can imagine that if a hundred individuals developed a sense of purpose and community around the creation of a new kitchen implement, they might also buy it when available for sale, and possibly stimulate purchase by others. Of course, these attributes of the Quirky system are the context in which Quirky innovates and so are assumed when evaluating the quality of an idea. But, to the extent that the engagement intrinsic to Quirky introduces some unique dynamic in the purchase process, our numerical results may be only illustrative and not apply to housewares firms that operate in different ways. Recent work by Bayus (forthcoming) has explored some of the idiosyncrasies of an idea crowdsourcing platform. We conjecture that, if anything, the Quirky system *limits* the variance in the exogenous factors E.

If so, we would expect that the variance in outcomes explained by the quality of the raw ideas would be *greater* for Quirky than for firms with more channels of distribution.

We have outcome measures for sales rates over eight months. Ideally we would observe outcomes over the entire lifecycle of the products. Although we observe that sales rates are fairly constant over the observation window, it is possible that if observed over many years, the outcomes might be different. For instance, it may be that early sales rates are more subject to exogenous factors, but that in the long run a truer indicator of commercial success would be revealed. To avoid this limitation empirically, one would either have to measure idea quality and then wait a very long time to conduct the analysis, or one would face the problem of retrospective evaluation of idea quality, which would be polluted by knowledge of sales outcomes, and by shifts in consumer preferences over the waiting period.

We attempted to match the profile of the purchase intent survey respondents with the profile of Quirky's customers. Inevitably that matching is inexact. With greater expense one might recruit a sample with a better fit to Quirky's customer population. Such an effort might result in purchase intent measures that are better predictors of outcomes than the ones we obtained. Indeed, it is possible that other methods might yield better estimates of idea quality than we were able to obtain. We observe that there is almost no scholarly research validating quality estimation techniques, and our work may provide some conceptual foundations and models for conducting such research in the future.

As mentioned in our Analysis and Results section, we directly observe only revenue, not profit. We have addressed this issue with a cost model, but actual observations of cost would reduce the noise in the costs and ensure that they were unbiased.

We have suggested that Quirky is trying to select the best ideas to develop, and we have considered the success of each idea individually. In other words, we have not considered portfolio effects, which would add a dependency among the successes of the projects. We ignored any portfolio effects for three reasons. First, the Quirky products are from such a broadly defined product space, that there seems to be little risk of cannibalization. Second, we see no evidence in their stated criteria—the five sub-scores of *community*, *staff*, *design*, *market*, and *viability*—that they are considering such effects. And third, the practical reason that collecting evaluations of incremental value to a given portfolio seems prohibitive. Conceptually, the question of portfolio composition is an interesting one, but does not appear to be a key issue in this setting.

It could be the case that experts are better at judging the relative promise of ideas in a narrow scope rather than a broad one. For example, Dahl and Moreau have subjects generate concepts as solutions for the commuting diner. Perhaps the ideas are easier to discriminate because they are easier to compare. However, the opposite could be true: given that the ideas are more similar, they could be harder to discriminate (Kornish and Ulrich 2011). Our work suggests that further investigation of the role of experts is warranted.

References

Alexa.com. 2011. Statistics Summary for Quirky.com. URL: http://www.alexa.com/siteinfo/quirky.com#, accessed May 5, 2011.

Arthur D. Little, Inc. 1921. On Making Silk Purses from Sows' Ears. URL: http://libraries.mit.edu/archives/exhibits/purse/, accessed January 18, 2012.

Astebro, T. 2003. The Return to Independent Invention: Evidence of Unrealistic Optimism, Risk Seeking or Skewness Loving? *The Economic Journal*, 113(484): 226-239.

Bayus, B. Crowdsourcing New Product Ideas Over Time: An Analysis of the Dell IdeaStorm Community, *Management Science*, forthcoming.

Clancy, K. J., P. C. Krieg, and M. M. Wolf. 2006. *Market New Products Successfully*. Lexington Books, Lanham, Maryland.

Chandy, R., B. Hopstaken, O. Narasimhan, and J. Prabhu. 2006. From Invention to Innovation: Conversion Ability in Product Development. *Journal of Marketing Research*, 43(3): 494-508.

Cronbach, L. J. 1951. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3): 297-334.

Dahan, E., A. J. Kim, A.W. Lo, T. Poggio, and N. Chan. 2011. Securities Trading of Concepts (STOC). Journal of Marketing Research. 48: 487-517.

Dahan, E. and H. Mendelson. 2001. An Extreme Value Model of Concept Testing. *Management Science* 47(1): 102-116.

Dahl, D. W. and P. Moreau. 2002. The Influence and Value of Analogical Thinking during New Product Ideation. *Journal of Marketing Research*. 39(1): 47-60.

Diehl, M. and W. Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*. 53(3): 497-509.

Eliashberg, J., S. K. Hui, and Z. J. Zhang. 2007. From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts. *Management Science*. 53: 881-893.

Faulkner, M. and D. R. Corkindale. 2009. Are Experts Better than Potential Users in Predicting the Uptake of an Innovation? Extending the Use of the Juster Scale. *Technological Forecasting & Social Change*, 76: 910-916.

Girotra K., C. Terwiesch, and K.T. Ulrich. 2010. Idea Generation and the Quality of the Best Idea. *Management Science*, 56(4): 591-605.

Goldenberg, J., D. R. Lehmann, and D. Mazursky. 2001. The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success. *Management Science*. 47: 69-84.

Goldenberg, J., D. Mazursky, and S. Solomon. 1999. Toward Identifying the Inventive Templates of New Products: A Channeled Ideation Approach. *Journal of Marketing Research*. 36: 200-210.

Golder, P. N., D. Mitra, and C. Moorman. What is Quality? An Integrative Framework for Processes and States. *Journal of Marketing*, forthcoming.

Griffin, A. and A. L. Page. 1993. An Interim Report of Measuring Product Development Success and Failure. *Journal of Product Innovation Management*, 10: 291-308.

Griffin, A. and A. L. Page. 1996. PDMA Success Measurement Project: Recommended Measures for Product Development Success and Failure. *Journal of Product Innovation Management*, 13:478-496.

Hauser, J., G. J. Tellis, and A. Griffin. 2006. Research on Innovation: A Review and Agenda for Marketing Science. *Marketing Science*. 25(6): 687-717.

Hoch, S. J. 1988. Who Do We Know: Predicting the Interests and Opinions of the American Consumer. *Journal of Consumer Research*. 15(3): 315-324.

Jamieson, L. F. and F. M. Bass. 1989. Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research*, 26(3): 336-345.

Kamakura, W. A., S. Basuroy, and P. Boatwright. 2006. Is Silence Golden? An Inquiry into the Meaning of Silence in Professional Product Evaluations. *Quantitative Marketing and Economics*. 4: 119-141.

Kornish, L. J. and K. T. Ulrich. 2011. Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas, *Management Science*, 57(1): 107-128.

Moore, W. L. 1982. Concept Testing. *Journal of Business Research*. 10: 279-294.

Morwitz, V. G., J. H. Steckel, and A. Gupta. 2007. When Do Purchase Intentions Predict Sales? *International Journal of Forecasting*, 23:347-364.

Ottum, B. D. 2005. Quantitative Market Research. Chapter 18 in K. B. Kahn (Ed.) *The PDMA Handbook of New Product Development, 2nd Edition*, John Wiley & Sons, Inc. Hoboken, New Jersey.

Ozer, M. 2009. The roles of product lead-users and product experts in new product evaluation. *Research Policy*, 1340-1349.

Ries, E. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* Crown Business.

Rogers, E. M. 1995. *Diffusion of Innovations*. The Free Press. New York.

Rust, R. T., C. Moorman, and P. R. Dickson. 2002. Getting Return on Quality: Revenue Expansion, Cost Reduction, or Both? *Journal of Marketing*. 66: 7-24.

Sackett, P. R. and H. Yang. 2000. Correction for Range Restriction: An Expanded Typology. *Journal of Applied Psychology*.85(1): 112-118.

Terwiesch, C. and K.T. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business Press.

Tetlock, P.E. 2005. *Expert Political Opinion, How Good is it? How Can we Know?* Princeton University Press. Princeton, NJ.

Ulrich, K. T. and S. D. Eppinger. 2011. *Product Design and Development*, 5th Ed. McGraw-Hill Higher Education.

Wooldridge, J. M. 1999. *Introductory Econometrics: A Modern Approach*. South Western.