The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis

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

Karl T. Ulrich
CIBC Professor of Entrepreneurship and e-Commerce
The Wharton School
University of Pennsylvania
500 Huntsman Hall
Philadelphia, PA 19104 USA
ulrich@wharton.upenn.edu

Version of August 2013

Abstract

How important is the original conception of an idea, the *raw* idea, to the success of an innovation? Classic work on diffusion of innovations by Rogers (1995) and more recent work on inventive templates by Goldenberg et al. (2001) have looked at the predictive power of qualitative characteristics of ideas. In this paper, we look at the question of whether raw ideas judged as *better* fare better in the market, and how strong that relationship is. We develop a novel data set from Quirky, a community-driven product development company that develops household consumer products. 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 two measures of idea quality: those from online consumer panelists and those from expert evaluators. First, we find that online consumer panels are a better way to determine what a "good" idea is than ratings by experts. Second, we find reliable predictions with samples as small as twenty consumers. Third, we find that there is a stronger predictive link between raw ideas and consumers' purchase intent of final product designs than there is between intention to purchase the final designs and market outcomes. Fourth, we find that the commercial importance of the raw idea is large, with a one-standard-deviation better idea translating to an approximately 50% increase in sales rate.

Keywords**:** innovation, new product development, raw ideas, market outcomes, new product evaluations

Acknowledgment

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 this work from Eric Bradlow, Cheryl Druehl, John Lynch, Page Moreau, Kamalini Ramdas, David Robertson, and participants at the 9th Annual Product and Service Innovation Conference and the 2012 Yale Center for Consumer Insights Conference. We thank Abigail Schneider for research assistance.

(This paper has been conditionally accepted in *Journal of Marketing Research*. Please cite the SSRN working paper or check for the published *JMR* article.)

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, "you can't make a silk purse out of a sow's ear." The *sow's ear hypothesis* is that there is a demonstrable connection between the quality of the raw idea and the success of the resultant innovation. 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 perspective that "ideas are overrated" because execution is what matters (Ries 2011). Bolstering *Midas* is the recognition that raw ideas are rarely truly novel (Kornish and Ulrich 2011, Johnson 2010): if the raw idea is not unique, then how can it be the seat of value?

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, the upper part of Figure 1 shows a visual depiction of a raw idea from our data set, along with the corresponding final design 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 could 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 question two ways. First, we analyze how much of the variation in outcomes in innovation is explained by variation in the quality of the raw ideas. Second, we look at how much the explained variation matters, in terms of quantitative impact on outcomes. Following on that question, we can also look at the best way to identify good ideas, given that idea quality is an elusive notion.

Insert Figure 1 about here.

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.

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: 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 are important. We find that the quality of raw ideas as estimated by commercially feasible techniques is a statistically significant predictor of market outcomes. The raw idea itself explains only a modest amount of variation in outcomes, but even the modest variation corresponds to an economically important impact on value. 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. We consistently find predictive power in a sample as small as twenty consumers.

In the paper, we first discuss prior research in related areas. Then we lay out the conceptual framework; discuss the data, the analysis, and the results; and conclude.

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 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, 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 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 design and then from final design to market outcome. 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.

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. As ideas progress through the new product development process, from idea generation to product launch, various activities and decisions contribute to value creation. In this study we examine both the role of the quality of the raw idea itself and the role of the development decisions that shape the ultimate product. In this section we consider how to think about the quality of the raw idea, how to think about the role of the design process, and how both the quality of the idea and the role of the design process can be analyzed together.

Defining the Quality of the Raw Idea

The true quality of a raw idea is a theoretical notion. Quality cannot be observed directly. We define the quality of the raw idea as a continuous variable reflecting the expected value of pursuing that idea, given the innovator's 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.

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. (2012) 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. (2012).

Resolution of Uncertainty in the Innovation Process

In most innovation processes, the innovator starts with an idea, and therefore, an estimate of value based largely on knowledge about the perceived quality of idea. Then, the development process resolves further uncertainty. Finally, the commercialization process is completed and the exogenous factors realized to reveal the value created. This typical sequence is shown in Figure

1. In our analysis, we examine the overall relationship between the quality of the raw idea and the value realized (the results of both Steps 1 and 2) and we study the two steps separately.

Overall relationship between raw idea and value. Once an idea is developed into a product and sold, a measure of value V can be tracked. Our central questions, then, are as follows. How much of the variation in V is explained by variation in I? How big is the impact on V from a better I? What is a good way to measure I?

Role of the development process. In our framework, we also consider an intermediate step, the result of the design and development process which creates the final design that is then sold to consumers. We would expect that some of the variation in innovation outcomes V would be explained not just by the quality of the raw idea I, but by what the innovator actually does with the idea. Armed with an estimate of the quality of the final design, D, we can decompose the analysis and look first at how much of the variance in D is explained by variance in I; and then at how much of the variance in V is explained by variance in D.

DATA

Our data set comprises raw ideas and final designs from a product development and commercialization company, independent evaluations of those ideas and designs, 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 typically contribute more than a hundred 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 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. 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. 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.

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 school.” (Alexa.com, 2013).

Ideas

Our data set comprises 160 products from the Quirky store. These products all appeared for sale in one of our two data collection periods, March to November 2011 and December 2012 to March 2013.[1] We retrieved ("scraped") from Quirky's site the description of the product in the store and images of the final designs. We collected sales figures on a regular basis during each data collection period.

We also retrieved the "raw idea" (text and images, if available) associated with each product in the store. There were 149 such raw ideas; some of the 160 products were developed from the same raw idea. The upper part of Figure 1 shows a visual depiction of a raw idea from our data set, along with the corresponding final design that was developed from the raw idea; Figure WA1 in the Web Appendix shows the accompanying text descriptions. We had two research assistants independently classify the products into categories, based on the room of the house for which the product was intended (bathroom, bedroom, garage/utility room, kid's room, kitchen, office). They agreed on 78% of the categories in their independent classifications; they then reconciled the discrepancies. The consensus of our research assistants showed that the 160 products comprised 14 bathroom products, 14 bedroom products, 38 garage/utility room products, 5 kids products, 49 kitchen products, and 40 office products.

Finally, we also generated a random sample of 100 raw ideas from all the idea 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.)

Measuring Value with Outcomes

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. 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. Figure 2 shows a sample of sales trajectories in the data. We use three approaches to address the varied launch dates. First, to normalize for the length of time in the store, we look at sales *rates*, units sold divided by the days in the store. Second, we use the sales trajectories to estimate projected units for each product. We made these projections with s-curve forecasts using the Bass model (Bass 1969, Srinivasan and Mason 1986). Using projected units relaxes some of the assumptions implicit in using sales rates, acknowledging that products are not offered for sale forever and that sales rates may vary over the life of a product. To solve the non-linear optimization problem for each product, we first did a grid search over the parameter space (the p, q, and m parameters of the Bass model) and then used SAS (proc nlin) to find the best fitting curve using our grid solution as a starting point. Third, for the models with sales rate as the dependent measure, we run variations including a control for the number of days in the store.

Because we do not know actual 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).

Further, to address potential endogeneity of price in the estimation, we use estimates of manufacturing costs. For the final designs, 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 manufacturing cost based on these factors, as prescribed by Ulrich and Eppinger (2011). Cost is modeled as a sum of materials costs, part processing costs, assembly costs, and transportation costs. Each of those elements is in turn modeled as a function of the product parameters. For instance, transportation costs are determined from the product package dimensions and prevailing freight costs between the manufacturing site (China) and the U.S. The correlation between the natural log of estimated cost and the natural log of price is 0.78.

In sum, we track the following value outcomes: sales rate, units sold, projected units (from the Bass model), and revenue. 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).

Insert Figure 2 about here.

Measuring Quality of the Raw Idea

We measure the quality of the raw idea two ways. The first measure is purchase intent of 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 of the 149 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 purchased from Qualtrics 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. The attention filter was formatted exactly like the concept descriptions in the survey, but in the place of the concept title was “Survey Reading Verification” and in place of the concept description was an instruction to select the leftmost option (the “definitely not”).

The 249 ideas were divided (randomly) into blocks of 49 or 50, with 29 or 30 ideas from the 149 ideas that made it into the store in each block, and 20 of the 100 random ideas in each block. Each panelist was assigned to a block and shown the ideas from that block in a random order. There were 1438 responses, with each of the five blocks rated by between 282 and 293 panelists. (The block design helps us compare the interrater reliability of these panelist with that of our experts. The unequal numbers came from random assignment, screening out, and incompletion.) We also collected self-reported information from the panelists on gender, age group, and employment status. Based on the self-reported information, the panel was 52.7% female. It was 13.1% ages 18-25, 24.8% ages 26-35, 22.8% ages 36-45, 20.2% ages 46-55. 14.5% ages 56-65, and 4.5% 66 and older. Students comprised 8.5% of the panelists, and 58.6% of the panelists reported working part- or full-time.

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). We present the results using this weighting, and the results hold when the weighted measures are logged.[2] Consistent with typical practice for testing raw ideas, for which design concepts are not yet developed, the product descriptions did not suggest prices.[3]

Our main analysis focuses on the ideas that made it into the store, but we also used the purchase intent results for the 100 random ideas. First, we use the standard deviation of purchase intent of the random ideas to estimate the impact of having a “one standard deviation better” idea. Second, we note that the standard deviation of purchase intent of the 149 ideas that made it into the store is very close to the standard deviation of purchase intent of the 100 random ideas: 0.0796 vs. 0.0841. That suggests that the results about the relationship between idea quality and outcome are not dramatically skewed by a restricted variance on idea quality in the set of selected ideas compared to the whole population.

The second measure of idea quality was ratings by experts. We used seven experts in consumer-products marketing and product development. They rated 98 of the store ideas plus fifty out of the set of 100 random ideas. (All seven experts saw the same 148 ideas.) The 98 ideas came from the initial phase of our data collection, and the 50 ideas were selected randomly from the set of 100 random ideas.[4] 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).

Experts were asked to rate the ideas on a scale of 0 to 10, based on anticipated units sold. The question to the experts 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 associated with raw ideas would not yet be determined, and we therefore directed the experts to assume that the "resulting products would be priced appropriately."

Measuring Quality of the Final Design

We measure the quality of the final design with purchase intent, using a procedure similar to that for the raw ideas. A different group of paid online panelists looked at product descriptions and images of the final designs, 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 53-54 final designs and faced an attention filter question. There were 363 responses, with each of the three blocks rated by between 113 and 129 panelists. Based on the self-reported information, the panel was 47.9% female. It was 7.7% ages 18-25, 13.5% ages 26-35, 16.5% ages 36-45, 27.5% ages 46-55. 27% ages 56-65, and 7.7% 66 and older. Students comprised 5.8% of the panelists, and 58.4% of the panelists reported working part- or full-time.

The descriptions of the final designs *did* include prices, as prices were set by Quirky as part of the final development process. For most of the products, the actual market prices were constant over the observation periods. For the products with price changes, we calculated a representative price, which we defined as the "modal quantity" price, or the price at which the largest quantity was sold. Of course we would expect a relationship between price and quantity, but we felt this approach was the most genuine way of capturing a representative price. This representative price was the price we used in our survey.

One of the challenges in finding a suitable data set for this study 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 4%) had heard of Quirky; we screened out those respondents.

In the measures described above, we have estimated the quality of the raw idea, the quality of the final design, and the market outcome, the essential ingredients for addressing our research questions. The descriptive statistics of our measures are summarized in Table 1 and the correlation table is shown in Table 2.

Insert Table 1 about here.

Insert Table 2 about here.

ANALYSIS AND RESULTS

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

We answer our central question about the importance of the raw idea by estimating the relationship between measures of idea quality and outcomes. Our data allow us to not only test that start-to-finish relationship, but to decompose the analysis into two steps: first, the relationship between raw idea quality and purchase intent of the final design, and second, the relationship between purchase intent of the final design and outcomes. See Figure 1.

To analyze the relationship between the raw idea *I* and the outcome, the arrow spanning both steps in Figure 1, we estimate Equations (1′) and (1″). In these equations, p represents a control variable for price and CAT represents a vector of dummy variables for product category. For concreteness, the dependent variable shown is Ln(Sales Rate), but we also perform the analysis for other outcome measures such as units sold and projected units.

$$\text{Ln(Sales Rate)} = \beta_0 + \beta_1 I + \beta_2 \text{Ln}(p) + \varepsilon \tag{1′}$$

$$\text{Ln(Sales Rate)} = \beta_0 + \beta_1 I + \beta_2 \text{Ln}(p) + \beta_3 \text{CAT} + \varepsilon \tag{1″}$$

Price is a natural control to use, and we expect a negative relationship between the price of the product and the quantity sold. To address the potential endogeneity of price in this model, we use an instrumental variable for it, estimated manufacturing costs. Our main results use that instrumental variable in a Two Stage Least Squares (2SLS) procedure. In 2SLS, we first estimate price as function of cost and the other exogenous variables (stage 1, included in the Web Appendix in Table WA1) and then use the estimated price in the regression for sales rate (stage 2, Table 3 and Table 4). The last lines of Table 3 and of Table 4 show the results of the Hausman test for endogeneity for each model. When that test shows significance, as it does for most of the models, the 2SLS estimates are preferred to Ordinary Least Squares (OLS). We include the analogous OLS analyses in Table A1 in the Appendix.

Product categories are another potential source of variation in outcomes—perhaps the degree of need is different across product categories—so we control for that too. Those model variations test whether category adds predictive power, even on top of price: e.g., do similarly-priced kitchen products tend to sell more (or less) than office products?

Throughout our analyses, we use the natural log of the sales rate and of the price: these two variables have values that extend over more than a factor of ten, with a long right tail. We report results with the non-logged versions of the idea quality measures. The results are not substantially different with the natural logs of those quantities too.

Purchase Intent

In this subsection, we present the results using purchase intent as the measure of idea quality. We see from the simple correlation in Table 2 that purchase intent of the raw idea is a statistically significant predictor of the natural log of sales rate (r=0.25): higher rated raw ideas predict higher performing products.

How much does the idea matter? We can answer that question two ways: variance explained and size of the effect on outcomes. The squared simple correlation and OLS partial R^2 measures[5] tell us that the raw idea accounts for between four and eight percent of the variance in the natural log of sales rate. Therefore, in terms of variance explained, the effect of the raw idea appears modest, but still statistically significant. In terms of impact, though, this small explained variance translates to a large effect on outcomes. The standard deviation of purchase intent of the raw idea in the *random* sample of ideas is 0.0841. In Models 1′ and 1″, the coefficients on purchase intent of the raw idea are 6.849 and 4.885, respectively. A one standard deviation change in purchase intent translates to a change in the natural log of sales rate estimated at 0.576

(=0.0841*6.849) for Model 1′ and 0.411 (=0.0841*4.885) for Model 1″. In other words, a one-sigma better idea corresponds to a 51% to 78% increase in sales rate. (The 51% = $e^{0.411}$ -1 and the 78% = $e^{0.576}$ -1.) Much of variance is unexplained, but even six percent of the variance in the natural log of sales rate corresponds to a quite large economic effect.

That analysis applies to the conjunction of the two steps in Figure 1. Now we turn our attention to each step individually. In Models A′ and A″ we examine Step 1 from Figure 1. We test whether the purchase intent of the raw idea predicts the purchase intent of the final design by estimating models analogous to the tests of Model 1, but with the dependent variable as the purchase intent of the final design.

$$\text{PI-Final Design} = \beta_0 + \beta_1 I + \beta_2 Ln(p) + \varepsilon \qquad (A')$$

$$\text{PI-Final Design} = \beta_0 + \beta_1 I + \beta_2 Ln(p) + \beta_3 CAT + \varepsilon \qquad (A'')$$

Again we find significance of the raw idea. The squared simple correlation and OLS partial R^2s of these models exceed 0.30. The purchase intent of the raw idea is a strong indicator of the purchase intent of the final design.

In Models 2′ and 2″ we examine Step 2 from Figure 1. We test whether the purchase intent of the final design predicts sales rate by estimating models analogous to the tests of Model 1, but with an independent variable of the purchase intent of the final design, D.

$$Ln(\text{Sales Rate}) = \beta_0 + \beta_1 D + \beta_2 Ln(p) + \varepsilon \qquad (2')$$

$$Ln(\text{Sales Rate}) = \beta_0 + \beta_1 D + \beta_2 Ln(p) + \beta_3 CAT + \varepsilon \qquad (2'')$$

In Models 2′ and 2″, the significance of the final design quality operates via price; negative and significant coefficients in the price estimation stage (Table WA1 in the Web Appendix) show that higher intent ratings are associated with lower prices. The survey for purchase intent of the final design included the price in the description of the product, and the correlation between purchase intent of the final design and the natural log of price is -0.45.

Comparing Step 1 to Step 2 from Figure 1 by looking at Models A″ and 2″, we see that the R^2 of the first step (0.52) is greater than that of the second step (0.31). The OLS partial R^2s also show the contrast between Step 1 and Step 2: controlling for price and category, purchase intent of the raw idea explains 38% of variance in purchase intent of the final design, but purchase intent of the final design only explains 3.7% of the variance in the natural log of sales rate.

Insert Table 3 about here.

Insert Table 4 about here.

These results are robust to other outcome measures and other approaches to controlling for different lengths of time that products were in the store. Tables WA2-WA4 in the Web Appendix show, respectively, the results when the outcome variable is units sold, Bass model projections on the whole data set (N=160), and Bass model projections for products in data set that have been in the store more than six months (N=109). Imposing a minimum sales history also increases the predictive power of purchase intent for sales rate.[6] For example, the R^2 for Model 1″ increases from 32% to 40% when the sample is limited to the 76 products that have been in the store for more than 450 days. The estimated coefficients on purchase intent, and thus the size of the effect on sales, also steadily increase with longer sales history. This pattern suggests that purchase intent improves as a predictor of market outcomes when the outcome measure is a longer term metric.

These strengthened results on subsamples of the data suggest that controlling for the days in the store will add explanatory power to the model. Indeed it does, and we provide the results of such an analysis in Table WA5 in the Web Appendix. The magnitude of the coefficients on the quality measures remain relatively stable compared to Table 3 and Table 4.

Our conclusions about purchase intent and outcomes are as follows. First, the quality of the raw idea as measured by purchase intent is a statistically significant predictor of outcomes. Second, even though the percent of variance explained is modest, the impact on outcomes is large. Third, there is a stronger predictive link between a better raw idea and better final design than there is between a better final design and better outcomes. And fourth, outcomes measured with more sales history are better explained by raw idea quality than those with less sales history.

Expert Ratings

Next we examine estimates of idea quality based on the ratings from our seven experts. We estimate versions of equations (1′) and (1″) and (A′) and (A″) with the average ratings of the experts as the measure of idea quality.

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 A2 in the Appendix. The correlation for the average of the experts with natural log of sales rate is higher than any individual expert. Poor agreement suggests that there will be poor interjudge reliability. In fact, common measures of interjudge reliability, Cronbach’s alpha (Cronbach 1951), which is 0.46 for these experts, and Krippendorff's alpha (Hayes and Krippendorff, 2007), which is 0.085, confirm that suggestion.

In spite of the low level of agreement and the poor reliability, the average of the experts' ratings is a statistically significant predictor of market outcomes, as shown in Models 3′ and 3″ in Table 3, and Models B′ and B″ in Table 4.

Examining Table 3, it would seem that the experts' predictive power for sales outcomes is comparable to that of the purchase intent of the idea for the consumer panel. Compare the R^2s in Models 1′ and 1″ (with purchase intent of the raw idea) of 0.16 and 0.32 to Models 3′ and 3″ (with experts) with 0.12 and 0.34. The increase in sales rate for a one-standard-deviation better idea is also similar, estimated at 48% [=exp(1.21*0.323)-1] for Model 3′ and 55% [=exp(1.21*0.364)-1] for Model 3″. (The corresponding values for the purchase intent of the raw idea, reported in the previous subsection, are 78% and 51%.)

What does not show in Table 3 is sensitivity to other dependent measures. The predictive power of the consumers' purchase intent ratings is largely robust to using units sold and projected units, but the same is not true of the experts' ratings. With these other dependent measures, the significance of the coefficient on the experts' average is marginal (with units sold, see Table WA2 in the Web Appendix) or not significant (with the projected units, see Tables WA3 and WA4 in the Web Appendix).

The contrast between experts' and consumers' predictive power is even more stark when we run the models only on the subset of ideas the experts rated, corresponding to N=109 products.[7] In those comparisons, Models 1′ and 1″ yield R^2s of 0.27 and 0.40 (those values are not shown in Table 3) vs. Models 3′ and 3″ with R^2s of 0.12 and 0.34 (as shown in Table 3). One explanation for the greatly increased predictive power on the smaller sample is that these products have a longer sales history, on average, than the full sample. As we discussed at the end of the previous subsection, restricting the analysis to products with longer sales histories in our data generally increases the predictive power of the models.

Another discrepancy in comparing experts and consumers is their number. We would like to know, "How big a set of consumers do we need to generate equivalent predictive power to our seven experts?" To identify the expert-equivalent sample size, we pull repeated subsamples from the consumer respondents, calculate the weighted purchase intent of the sample, and run the regression models. In the results reported below, we iterated 500 times, that is, pulled 500 subsamples for each sample size we tried. A sample of 10 consumer respondents means that we pulled 10 respondents from each of the five blocks, so that each idea was rated exactly 10 times.

The left panel of Figure 3 shows the average R^2 across the 500 trials for eight different sample sizes (1, 4, 7, 10, 15, 20, 30, and 50) for Model 1′ and Model 1″ with the set of 109 products whose ideas were rated by experts. With a sample size of four, the average R^2 for Model 1″ is 0.33. That is directly comparable to the experts' 0.34 (Model 3″ in Table 3).

The right panel of Figure 3 shows the percentage of the 500 iterations that yielded a significant coefficient on the purchase intent of the raw idea in Model 1″. In 52% of the 500 trials, the purchase intent calculated from only *four* consumers significantly predicts the outcome measure (natural log of sales rate) at $p < 0.05$ (and in 28% of the 500 at $p < 0.01$). For as few as 15 consumers, we see more than a 90% chance of having their collective voice be a significant predictor ($p < 0.05$) of outcomes. With any 20 consumers from our sample, we were virtually guaranteed to see significance at the $p < 0.05$ level.

Insert Figure 3 about here.

We have two conclusions from these data. First, asking consumers to state purchase intention of raw ideas predicts market outcomes better than asking experts to predict sales. Consumers look even more attractive when considering the higher cost of enlisting experts. Second, the consumer results are extremely robust, in terms of likelihood of significant prediction and in terms of explanatory power, even at small sample sizes. The answer to the question of “about how many consumers are equivalent predictors to our seven experts?” is about 4-20: with only four consumers, on average, the predictive power is the same as the expert group, and with twenty, the predictive power is greater than that of experts, with near certainty of significance of the ratings. This type of consumer research is not typically used at such an early stage, partly, perhaps, for reasons of secrecy. Our evidence shows that market wisdom is embedded in even small crowds (cf. Surowiecki 2004).[8] Ideas can reliably be vetted early, cheaply and with limited dissemination.

The model of Gross (1972) shows that the reliability, or signal-to-noise ratio, of the screening instrument is an important determinant of value in a creative process. The potency of small samples of consumers suggests that the noise in a consumer purchase intent survey is relatively low, making these types of surveys a useful tool in creating value.

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: ideas are rarely novel, but truly great execution is rare, so the team and its ability to execute is what matters. At the root of this belief 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. These results suggest that higher fidelity screening in the earlier stages may be worth the investment. Our surveys yield predictive information about market outcomes, but other approaches warrant further research. Approaches that simulate markets, like Dahan et al.'s (2011) STOC, or if the prototyping economics allow, test markets with actual purchases, will help one determine whether one possesses silk or pigs' ears in the first place.[9]

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. Collecting this type of information is currently so inexpensive that not doing so seems foolish.

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.

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. 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. Recent work by

Bayus (2013) 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. 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.

In this study, our data do not allow us to deconstruct all the drivers of value. In particular, we don't know whether better ideas attract more talented developers, or better ideas are promoted more heavily, and therefore we can't say whether those factors are the reasons for better performance. On the one hand, this is a gap in our understanding of value creation in innovation. On the other hand, we can say that in the setting we studied, if these forces were at work, overall they did combine in a way so that ideas rated as better by consumers and experts had better market outcomes.

We note that we treat price differently in the two steps (as shows in Figure 1): in surveying consumers and experts about the raw ideas, we do not include prices, yet in surveying consumers about the final designs, we did include price. We chose that information structure because it represents what companies would actually know at each step. In a future study, we could further control for the effect of price in the comparison of the steps by collecting a version of the purchase intent of the raw idea which included the prices, because the stated prices do seem to have a strong negative effect on stated purchase intent.

We attempted to match approximately the profile of the purchase intent survey respondents with the profile of Quirky's customers. Inevitably that matching is inexact. A sample better fit to Quirky's customer population might result in purchase intent measures that are better predictors of outcomes than the ones we obtained.

As mentioned in our Data 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 also used several dependent measures because we don't have the ideal measure of long term incremental profit.

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 areas of *community*, *staff*, *design*, *market*, and *viability*—that they are considering such effects. And third, measuring incremental value to a given portfolio appears to be prohibitive practically. 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 (Kornish and Ulrich 2011), they could be harder to discriminate. Our work suggests that further investigation of the role of experts is warranted.

REFERENCES

Alexa.com. 2013. Statistics Summary for Quirky.com. URL: http://www.alexa.com/siteinfo/quirky.com#, accessed March 28, 2013.

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.

Bass, F. M. 1969. A New Product Growth Model for Consumer Durables, *Management Science*, 15 (1): 251-27.

Bayus, B. 2013. Crowdsourcing New Product Ideas Over Time: An Analysis of the Dell IdeaStorm Community, *Management Science*, 59(1): 226-244.

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. 2012. What is Quality? An Integrative Framework for Processes and States. *Journal of Marketing*, 76(4): 1-23.

Griffin, A. and J.R. Hauser. 1993. The Voice of the Customer. *Marketing Science*. 12(1): 1-27.

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.

Gross, I. 1972. The Creative Aspects of Advertising. *Sloan Management Review*, 14(1): 83-109.

Haley, R. I. and P. B. Case. 1979. Testing Thirteen Attitude Scales for Agreement and Brand Discrimination. *Journal of Marketing*. 43(4): 20-32.

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.

Hayes, A. F. and K. Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures, 1(1): 77–89.

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.

Johnson, S. 2010. *Where Good Ideas Come From: The Natural History of Innovation*. Riverhead Books, New York.

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.

Srinivasan, V.and C. H. Mason 1986. Nonlinear Least Squares Estimation of New Product Diffusion Models. *Marketing Science*. 5(2): 169-178.

Surowiecki, J. 2004. The Wisdom of Crowds. Random House, Inc. New York.

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

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

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

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

[1] We resumed data collection in December 2012 based on requests from reviewers.

[2] Haley and Case (1979) document the non-linear relationship between sales and intent.

[3] 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).

[4] The 109 products resulting from ideas rated by the experts comprised 6 bathroom products, 10 bedroom products, 25 garage/utility room products, 5 kids products, 31 kitchen products, and 32 office products.

[5] The R^2s from the 2SLS procedure are not, strictly speaking, interpretable as percentage of variance explained, although they still have the form of such a measure, 1-sum of squared errors/total sum of squares (Wooldridge, 1999). The R^2s from both 2SLS and OLS are close in all of our models, with the 2SLS R^2s slightly lower; the OLS R^2s and the OLS partial R^2s for the quality variable are reported in Table A1.

[6] We also note that the analysis of a sample restricted by sufficient length of sales history does not show evidence of price endogeneity.

[7] To elaborate on the discrepancy in the samples for the experts and the consumers: the difference came from the timing and progress of the study. Our first round of data collection was in 2011, and our experts only rated the ideas in the first round. When we did a second round of data collection to update the sales figures and collect consumer ratings of all the ideas (including ones added since our first collection), we did not re-collect from the experts for two reasons. First, unlike the consumer panel, which is managed by a vendor (Qualtrics) from which we can purchase respondents, the experts are specific individuals who had already expended considerable time and effort for our study. Second, with the data from the first round, experts' ratings were not predictive of market outcomes, so that did not seem a promising use of their time and goodwill.

[8] We note that the specific numbers are reminiscent of the result in Griffin and Hauser (1993) that a subset of 20 customers name over 90% of the total needs identified.

[9] We thank an anonymous reviewer for this suggestion.

TABLES

Table 1: Descriptive Statistics.

	Mean	Median	Standard Deviation	N
Price (USD)	30.12	24.99	24.83	160
Units Sold	17,257	490	66,124	160
Projected Units	192,834	705	1,964,559	160
Sales Rate (units/day)	37.80	3.27	91.03	160
Revenue (USD)	334,749	12,815	1,519,439	160
Days in Store	462.1	400.0	360.3	160
Cost (USD)	6.29	4.35	6.44	160
Purchase Intent Raw Idea (0-1), Developed Ideas	0.45	0.46	0.08	149
Purchase Intent Raw Idea (0-1), Random Ideas	0.40	0.40	0.08	100
Purchase Intent Final Design (0-1)	0.28	0.27	0.08	160
Expert Rating (0-10), Developed Ideas	4.26	4.29	1.22	98
Expert Rating (0-10), Random Ideas	3.37	3.21	1.21	50

Table 2: Correlation Matrix. Correlation coefficients for the variables for ideas that were developed and sold in the store. Correlations with average expert ratings contain only 109 observations, compared to 160 for the others. (*p< 0.10, **p<0.05, ***p<0.01)

	Ln(Price)	Ln(Units Sold)	Ln(Proj. Units)	Ln(Sales Rate)	Ln(Rev.)	Ln(Days in Store)	Ln(Cost)	PI-Raw Idea	PI-Final Design	Ave. Expert Rating
Ln(Price)	-									
Ln(Units Sold)	-0.35***	-								
Ln(Projected Units)	-0.29***	0.89***	-							
Ln(Sales Rate)	-0.35***	0.85***	0.73***	-						
Ln(Revenue)	-0.02	0.94***	0.84***	0.78***	-					
Ln(Days in Store)	-0.01	0.31***	0.32***	-0.24***	0.33***	-				
Ln(Cost)	0.78***	-0.44***	-0.38***	-0.40***	-0.19**	-0.08	-			
PI-Raw Idea	0.02	0.23***	0.17**	0.25***	0.26***	-0.03	0.06	-		
PI-Final Design	-0.45***	0.30***	0.19**	0.36***	0.16**	-0.09	-0.21***	0.55***	-	
Ave. Expert Rating	-0.09	0.18*	0.14	0.24**	0.17*	-0.14	0.04	0.49***	0.31***	-

Table 3: Sales Models. Results of second stage of 2SLS regressions for the variants of Models 1, 2, and 3 as shown in Figure 1. Standard errors in parentheses.
(*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Ln(Sales Rate)

	Raw Idea→Sales		Final Design→Sales		Experts→Sales	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	3.322*** (1.228)	3.501*** (1.231)	4.910*** (1.434)	4.247*** (1.371)	4.507*** (1.423)	2.975* (1.594)
Estimated Ln(Price)	-1.573*** (0.276)	-1.330*** (0.246)	-1.401*** (0.313)	-1.155*** (0.274)	-1.506*** (0.360)	-1.212*** (0.314)
PI-Raw Idea	6.849*** (1.901)	4.885*** (1.810)				
PI-Final Design			3.727 (2.280)	3.591* (2.115)		
Ave. Expert Rating					0.323** (0.154)	0.364** (0.141)
Controls for Category	No	Yes	No	Yes	No	Yes
R^2	0.16	0.32	0.13	0.31	0.12	0.34
N	160	160	160	160	109	109

Hausman Test: Test of endogeneity.

	Model 1′	Model 1″	Model 2′	Model 2″	Model 3′	Model 3″
First Stage Residuals	1.344*** (0.423)	0.645 (0.415)	1.845*** (0.474)	0.954** (0.474)	1.441** (0.568)	0.725 (0.549)

Table 4: Final Design Models. Results of second stage of 2SLS regressions for the variants of Models A and B as shown in Figure 1. Standard errors in parentheses. (*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Purchase Intent of Final Design

	Raw Idea→Final Design		**Experts→Final Design**	
	Model A′	**Model A″**	**Model B′**	**Model B″**
Constant	0.129*** (0.037)	0.138*** (0.040)	0.296*** (0.050)	0.282*** (0.060)
Estimated Ln(Price)	-0.035*** (0.008)	-0.036*** (0.008)	-0.033** (0.013)	-0.029** (0.012)
PI-Raw Idea	0.560*** (0.057)	0.551*** (0.059)		
Ave. Expert Rating			0.018*** (0.005)	0.020*** (0.005)
Controls for Category	No	Yes	No	Yes
R^2	0.49	0.52	0.23	0.33
N	160	160	109	109

Hausman Test: Test of endogeneity.

First Stage Residuals	-0.045*** (0.013)	-0.049*** (0.013)	-0.033 (0.020)	-0.050** (0.020)

FIGURES

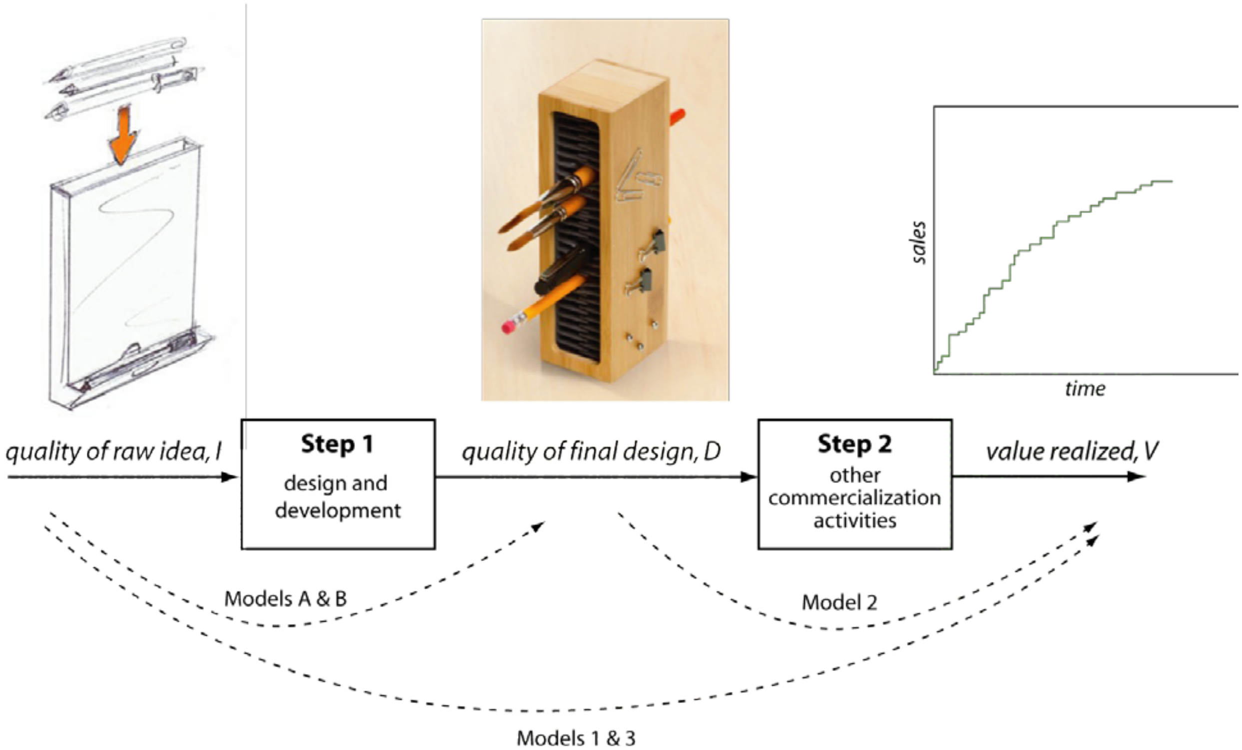


Figure 1: Two Steps of Value Creation. We analyze the overall relationship between raw idea and value (Models 1 & 3 in our empirical analysis). We also decompose that relationship into two parts: from raw idea to final design (Models A & B) and from final design to value realized (Model 2). We denote models with controls for price with a prime, e.g., Model 1′ and with controls for both price and categories with a double prime, e.g., Model 1″.

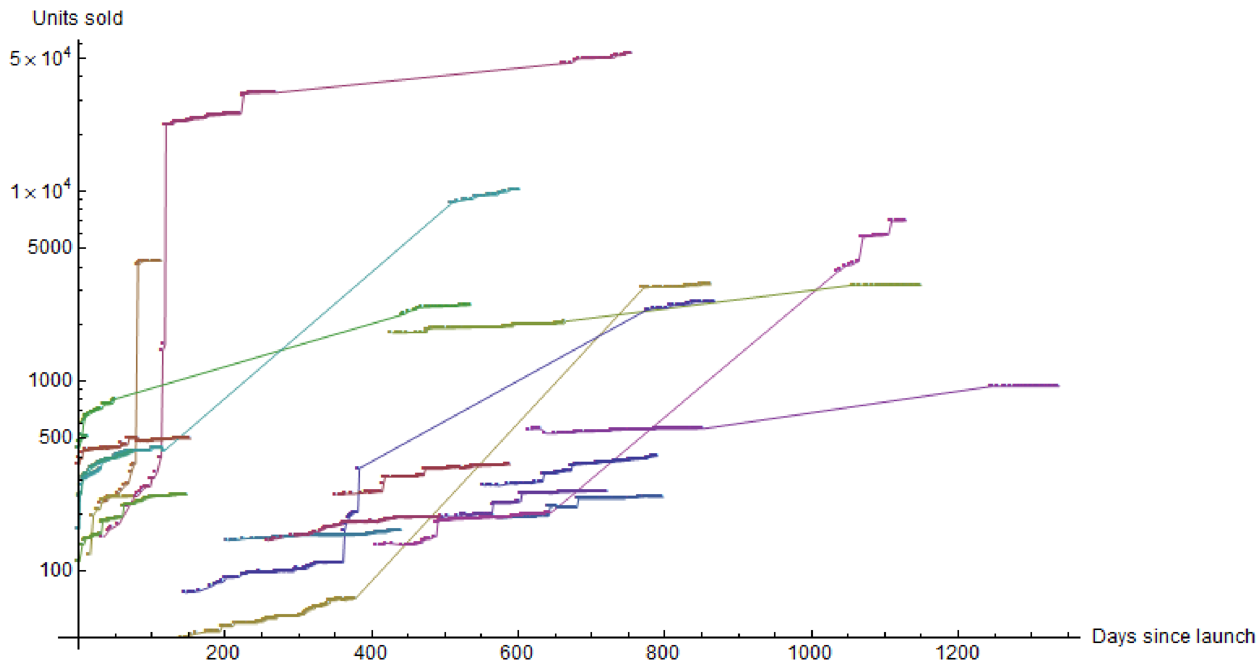


Figure 2: Sales Trajectories. Sales trajectories for 20 of the products in the sample. Drops in units sold reflect returns. The trajectories shown are the sales over the observation periods from March 25, 2011 to November 17, 2011 and December 14, 2012 to March 16, 2013. The northeasterly straight lines interpolate in the non-collection interval; vertical steps represent large orders, likely from retail partners.

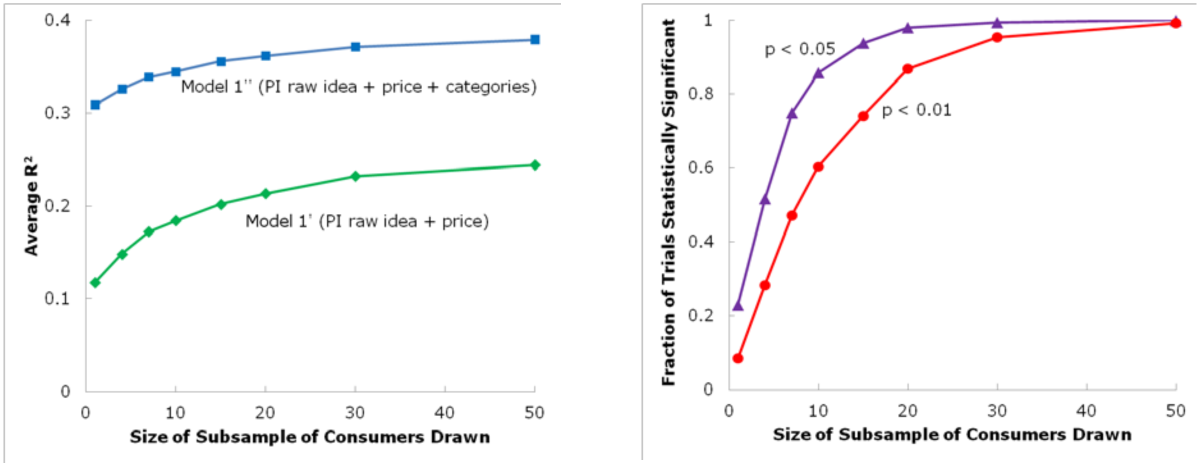


Figure 3: Consumer Subsamples. The left panel shows the average R^2 over 500 iterations as a function of the size of the consumer subsamples drawn. The lines are Model 1′, which includes the purchase intent of the raw idea and the price (bottom) and Model 1″, which adds category controls to Model 1′. The right panel shows the percent of the 500 iterations which showed a significant predictive effect of the purchase intent of the raw idea in Model 1″, as a function of the size of the consumer subsamples drawn. The top line is for significance at the p<0.05 level and the bottom line at p<0.01.

APPENDIX

Table A1: OLS Sales and Final Design Models. Results of OLS regressions for the variants of Models 1, 2, 3, A, and B as shown in Figure 1. Standard errors in parentheses. (*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Ln(Sales Rate)

	Raw Idea→Sales		**Final Design→Sales**		**Experts→Sales**	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	1.726 (1.088)	2.838** (1.149)	1.956* (1.150)	2.928** (1.191)	2.674** (1.188)	2.070 (1.431)
Ln(Price)	-1.052*** (0.212)	-1.108*** (0.198)	-0.703*** (0.240)	-0.847*** (0.224)	-0.974*** (0.281)	-0.978*** (0.258)
PI-Raw Idea	6.754*** (1.865)	4.839*** (1.803)				
PI-Final Design			6.469*** (2.096)	4.856** (2.003)		
Ave. Expert Rating					0.350** (0.151)	0.378*** (0.140)
Controls for Category	No	Yes	No	Yes	No	Yes
N	160	160	160	160	109	109
R^2	0.19	0.33	0.17	0.32	0.15	0.34
Adj. R^2	0.18	0.29	0.16	0.29	0.14	0.30
Partial R^2 (Idea Quality)	0.077	0.045	0.057	0.037	0.048	0.067

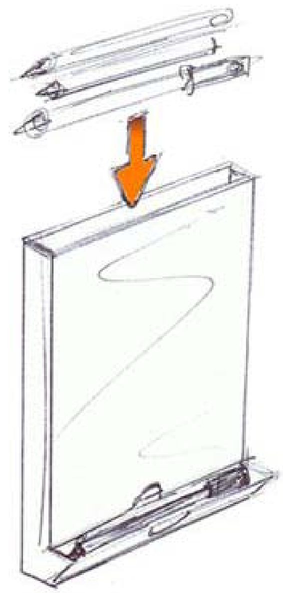
Dependent Variable: Purchase Intent of Final Design

	Raw Idea→Final Design		**Experts→Final Design**	
	Model A′	**Model A″**	**Model B′**	**Model B″**
Constant	0.183*** (0.033)	0.189*** (0.037)	0.338*** (0.042)	0.344*** (0.053)
Ln(Price)	-0.053*** (0.006)	-0.053*** (0.006)	-0.045*** (0.010)	-0.045*** (0.010)
PI-Raw Idea	0.563*** (0.056)	0.554*** (0.058)		
Ave. Expert Rating			0.017*** (0.005)	0.019*** (0.005)
Controls for Category	No	Yes	No	Yes
N	160	160	109	109
R^2	0.51	0.54	0.24	0.35
Adj. R^2	0.51	0.52	0.23	0.30
Partial R^2 (Idea Quality)	0.39	0.38	0.097	0.11

Table A2: Correlation Matrix for Experts. Correlation matrix of expert ratings on 148 raw ideas plus correlations with the natural log of sales rate for the 109 ideas that were launched. Experts rated 98 of the raw ideas that were developed into products and 50 of the randomly selected products. All seven experts rated the same 148 ideas.

	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	-
Ln(Sales Rate)	-0.016	0.185	-0.036	0.084	0.216	0.160	0.111	0.237

WEB APPENDIX



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 extending out of the top, or stand it upright with the items extending out of the ends. It's your call--choose the arrangement that suits you best! Hidden magnets below the bamboo surface hold paper clips and other small magnetic items.
Features:
- Rubber extrusions that act as magic fingers to hold your items in place.
- Bamboo exterior makes for an appealing addition to your desktop decor.
- Hidden magnets for extra storage of small magnetic items. With Pen Zen, your desk clutter can finally find inner peace.
Dimensions: 7.25in x 2.5in x 2.0in.

Price: $19.99

Figure WA1: Example raw idea (left) and final design (right).

Table WA1: Results of first (price estimation) stage of 2SLS regressions for the variants of Models 1, 2, 3, A, and B as shown in Figure 1. Standard errors in parentheses.
(*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Ln(Price)

	Cost, Raw Idea→Price		**Cost, Final Design →Price**		**Cost, Experts→Price**	
Model(s)	**1′ and A′**	**1″ and A″**	**2′**	**2″**	**3′ and B′**	**3″ and B″**
Constant	2.465*** (0.208)	2.568*** (0.238)	3.15*** (0.129)	3.140*** (0.159)	2.722*** (0.154)	2.928*** (0.234)
Ln(Cost)	0.564*** (0.036)	0.602*** (0.035)	0.518*** (0.032)	0.551*** (0.031)	0.547*** (0.041)	0.576*** (0.039)
PI-Raw Idea	-0.222 (0.427)	-0.653 (0.436)				
PI-Final Design			-2.619*** (0.391)	-2.797*** (0.374)		
Ave. Expert Rating					-0.068** (0.032)	-0.083*** (0.031)
Controls for Category	No	Yes	No	Yes	No	Yes

Table WA2: Results of second stage of 2SLS regressions for the variants of Models 1, 2, and 3 as shown in Figure 1, using *Units Sold* as dependent variable. Standard errors in parentheses.
(*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Units Sold

	Raw Idea→Sales		Final Design→Sales		Experts→Sales	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	9.676*** (1.277)	9.068*** (1.250)	11.927*** (1.500)	10.419*** (1.401)	11.603*** (1.512)	9.894*** (1.710)
Estimated Ln(Price)	-1.743*** (0.288)	-1.491*** (0.249)	-1.641*** (0.327)	-1.356*** (0.280)	-1.672*** (0.382)	-1.346*** (0.337)
PI-Raw Idea	6.498*** (1.977)	4.984*** (1.837)				
PI-Final Design			1.535 (2.385)	2.089 (2.161)		
Ave. Expert Rating					0.239 (0.163)	0.269* (0.151)
Controls for Category	No	Yes	No	Yes	No	Yes
R^2	0.13	0.33	0.086	0.31	0.12	0.32
N	160	160	160	160	109	109

Hausman Test: Test of endogeneity.

First Stage Residuals	1.752*** (0.427)	1.013** (0.414)	2.222*** (0.483)	1.304*** (0.477)	1.501** (0.604)	0.756 (0.590)

Table WA3: Results of second stage of 2SLS regressions for the variants of Models 1, 2, and 3 as shown in Figure 1, using *Projected Units* as dependent variable. Standard errors in parentheses.
(*p< 0.10, **p<0.05, ***p<0.01)

Dependent Variable: Projected Units—All 160 Products

	Raw Idea→Sales		Final Design→Sales		Experts→Sales	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	10.377*** (1.462)	9.639*** (1.480)	13.086*** (1.704)	11.373*** (1.650)	11.723*** (1.698)	9.515*** (1.998)
Estimated Ln(Price)	-1.687*** (0.329)	-1.434*** (0.295)	-1.671*** (0.372)	-1.378*** (0.330)	-1.508*** (0.429)	-1.184*** (0.394)
PI-Raw Idea	5.391** (2.263)	3.882* (2.176)				
PI-Final Design			-0.988 (2.710)	-0.194 (2.546)		
Ave. Expert Rating					0.196 (0.184)	0.236 (0.177)
Controls for Category	No	Yes	No	Yes	No	Yes
R^2	0.077	0.24	0.046	0.23	0.083	0.23
N	160	160	160	160	109	109

Hausman Test: Test of endogeneity.

First Stage Residuals	1.769*** (0.498)	1.019** (0.494)	2.090*** (0.567)	1.153** (0.571)	1.272* (0.692)	0.542 (0.694)

Table WA4: Results of 2SLS regressions for the variants of Models 1, 2, and 3 as shown in Figure 1, using *Projected Units* as dependent variable, for products in the store longer than 180 days. Stage 1, estimating Ln(Price) is included because the sample is different from that of Table WA1. Standard errors in parentheses. (*p< 0.10, **p<0.05, ***p<0.01)

Stage 1 Dependent Variable: Ln(Price)

	Cost, Raw Idea→Price		Cost, Final Design →Price		Cost, Experts→Price	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	2.253*** (0.260)	2.460*** (0.312)	2.984*** (0.170)	3.144*** (0.223)	2.863*** (0.165)	3.243*** (0.253)
Ln(Cost)	0.571*** (0.043)	0.606*** (0.043)	0.535*** (0.041)	0.563** (0.039)	0.540*** (0.042)	0.562*** (0.040)
PI-Raw Idea	0.247 (0.553)	-0.235 (0.563)				
PI-Final Design			-2.084*** (0.524)	-2.414*** (0.515)		
Ave. Expert Rating					-0.097*** (0.034)	-0.110*** (0.033)
Controls for Category	No	Yes	No	Yes	No	Yes

Stage 2 Dependent Variable: Projected Units—For Products In Store > 180 days

	Raw Idea→Sales		Final Design→Sales		Experts→Sales	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	9.280*** (1.665)	8.612*** (1.877)	10.763*** (1.907)	9.458*** (2.091)	12.511*** (1.896)	10.352*** (2.371)
Estimated Ln(Price)	-1.648*** (0.375)	-1.409*** (0.354)	-1.388*** (0.417)	-1.159*** (0.392)	-1.532*** (0.460)	-1.216*** (0.437)
PI-Raw Idea	8.378*** (2.781)	6.844** (2.833)				
PI-Final Design			5.677* (3.173)	5.532* (3.235)		
Ave. Expert Rating					0.054 (0.206)	0.176 (0.206)
Controls for Category	No	Yes	No	Yes	No	Yes
R^2	0.19	0.27	0.17	0.26	0.093	0.21
N	109	109	109	109	97	97

Hausman Test: Test of endogeneity.

First Stage Residuals	0.880 (0.609)	0.298 (0.617)	1.192* (0.661)	0.496 (0.683)	1.037 (0.755)	0.358 (0.784)

Table WA5: Results of 2SLS regressions for the variants of Models 1, 2, and 3 as shown in Figure 1, using *Natural Log of Sales Rate* as dependent variable, controlling for the natural log of days in the store. Standard errors in parentheses.
(*p< 0.10, **p<0.05, ***p<0.01)

Stage 1 Dependent Variable: Ln(Price)

	Cost, Raw Idea→Price		**Cost, Final Design →Price**		**Cost, Experts→Price**	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	2.265*** (0.274)	2.423*** (0.282)	3.05*** (0.212)	3.058*** (0.213)	2.992*** (0.431)	3.160*** (0.449)
Ln(Cost)	0.567*** (0.036)	0.605*** (0.036)	0.520*** (0.033)	0.552*** (0.031)	0.544*** (0.041)	0.573*** (0.040)
Ln(Days)	0.034 (0.030)	0.027 (0.030)	0.015 (0.027)	0.015 (0.026)	-0.041 (0.061)	-0.036 (0.059)
PI-Raw Idea	-0.211 (0.437)	-0.658 (0.436)				
PI-Final Design			-2.595*** (0.395)	-2.784*** (0.376)		
Ave. Expert Rating					-0.071** (0.032)	-0.086*** (0.031)
Controls for Category	No	Yes	No	Yes	No	Yes

Stage 2 Dependent Variable: Ln(Sales Rate)

	Raw Idea→Sales		**Final Design→Sales**		**Experts→Sales**	
	Model 1′	**Model 1″**	**Model 2′**	**Model 2″**	**Model 3′**	**Model 3″**
Constant	6.038*** (1.410)	6.189*** (1.326)	7.872*** (1.651)	7.147*** a (1.488)	4.561* (2.593)	2.865 (2.502)
Estimated Ln(Price)	-1.646*** (0.269)	-1.408*** (0.234)	-1.503*** (0.307)	-1.249*** (0.263)	-1.507*** (0.367)	-1.210*** (0.319)
Ln(Days)	-0.427*** (0.127)	-0.483*** (0.117)	-0.422*** (0.131)	-0.470*** (0.119)	-0.008 (0.296)	0.016 (0.271)
PI-Raw Idea	6.699*** (1.852)	4.933*** (1.728)				
PI-Final Design			2.802 (2.245)	2.885 (2.033)		
Ave. Expert Rating					0.322** (0.157)	0.365** (0.144)
Controls for Category	No	Yes	No	Yes	No	Yes
R^2	0.21	0.38	0.17	0.37	0.12	0.34
N	160	160	160	160	109	109

Hausman Test: Test of endogeneity.

First Stage Residuals	1.531*** (0.405)	0.827** (0.393)	2.008*** (0.456)	1.121** (0.450)	1.447** (0.574)	0.727 (0.555)