2012-07-06

Development and Initial Validation of an Innovation Assessment

Jacob D. Wheadon
Brigham Young University - Provo

Development and Initial Validation of an Innovation Assessment

Jacob D. Wheadon

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Geoffrey A. Wright, Chair
Richard E. West
Paul T. Skaggs

School of Technology

Brigham Young University

August 2012

ABSTRACT

Development and Initial Validation of an Innovation Assessment

Jacob D. Wheadon
School of Technology
Master of Science

In the past two decades, there has been an increased demand for more innovative individuals and organizations. In response to this need, a number of groups have begun to teach innovation courses to improve people's innovation skills. While many of these groups report success in helping people become more innovative, there is no way to test the effectiveness of the innovation courses.

This study describes the development and initial validation of an innovation test instrument. It demonstrates how the author identified the content domain of the test and created test items. Then it describes initial validation testing of the instrument.

This study found that this assessment is a good first step in creating an innovation assessment that covers more of the full process of innovation than previous tests. It still needs further validation and improvement to make strong claims about its ability to determine the effectiveness of an innovation course.

Keywords: innovation, assessment, validity, creativity

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 The Need for Innovation

In industry and education, there is an increasing push for organizations and individuals to be more innovative (Wagner, 2010; Fagerberg, 1999). Rapid technological change has created the need for organizations and individuals to adapt quickly (Christensen and Eyring, 2011). Christensen (1997) described how disruptive innovations fundamentally change markets and require new ways of thinking for organizations to adapt and survive. He described how individuals in organizations need to think differently in order to compete in today's marketplace. Because of the rapid rate of technological change that is occurring today, disruptive innovations are changing markets even faster than in the past. This has led to a greater need for people to cultivate innovation skills.

Innovation skills are also needed to create job growth. Drucker (1985) showed that innovation has been the leading source of job creation in the United States over the last century. He called for organizations and individuals to focus their efforts on creating new value in society, both for their own good, and for the good of society in general. These calls have been echoed by politicians (Obama, 2011), economists (Friedman and Mandelbaum, 2011), and educators (Wagner, 2010).

1.2 **The Need to Teach Innovation**

Many of these calls for increased innovation have mentioned the need for schools to teach students to be more innovative (Friedman and Mandelbaum, 2011; Wagner, 2010; Wagner 2012). They have said that for American students to remain competitive in a global market, and be able to adapt to a constantly shifting playing field, they need to become innovators. Schools need to teach students the skills and behaviors of great innovators (Wagner 2008).

In a recent study (Dyer et al., 2011), researchers found the common behaviors that many of today's leading innovators share. By studying innovators' behaviors, these researchers found that people who want to be better innovators can learn and practice behaviors that will help them create innovations. This is important because many creativity researchers have focused on links between creativity and personality (Hurt et al., 1977). Although these researchers found correlations between creativity and personality, these connections are not helpful to those hoping to increase a person's ability to innovate. By identifying behaviors, Dyer et al. (2011) give educators a set of teachable skills that students can learn to perform. They claimed that while some people might have a natural propensity for innovation, anyone can learn to be more innovative.

With the knowledge that innovation can be taught, some schools, consulting firms, and corporations have begun teaching innovation. Well-known examples include the Stanford D-School, IDEO, and Innosight. These groups have reported the great value and impact of their teaching about innovation (Stanford, 2011; IDEO, 2011; Innosight, 2011). This has led to many other schools and groups to attempt to teach innovation as well.

In order to keep up with the demand for innovation education, educators at Brigham Young University have developed a course focused on teaching innovation. The course, titled

the Innovation Bootcamp teaches BYU Technology and Engineering students the behaviors and processes of innovation that have been identified in past literature (Howell et al., 2011). At the Innovation Bootcamp, students learn tools that help them work through the five parts of the innovation model (as defined by the Innovation Bootcamp curriculum): idea finding, idea shaping, idea defining, idea refining, and idea communicating.

1.3 The Need to Assess Innovation Teaching

Using this model, educators have taught the Innovation Bootcamp since 2008. They have done preliminary studies (Howell et al., 2011; Wright et al., 2010) and feel confident that the course is having a positive impact on the innovation skills of the students, even though they did not have a test to formatively evaluate the impact the course was having on students' ability to innovate. Consequently, they need an assessment of students' innovation skills that can be used as a pre- and post- test to see if a student is more innovative as a result of participating in the Innovation Bootcamp. Having an innovation test would be very useful for improving teaching in this particular course, and it is hoped that such a test will have value for anyone seeking to teach innovation.

1.4 Current Innovation Assessments

In an attempt to address this need, Tyler Lewis' master's thesis (2011) reviewed existing innovation and creativity tests and relevant literature. His study found that existing test instruments were lacking in two major areas. The first is that existing tests do not cover the whole process of innovation – focusing only on either creativity or implementation. He found that creativity-centric tests measure divergent thinking, while existing innovation tests focus primarily on convergent thinking. Lewis states that this is problematic because innovation

involves both divergent and convergent thinking. He also suggested that the other issue of the innovation tests was that they only measured the performance of a product, team, or organization, and did not account for, or measure, the abilities of an individual. This does not allow educators to see how their instruction changes a student's ability to innovate. In order to meet the needs of the BYU Innovation Bootcamp and other innovation educators, a test that measures an individual's ability to do activities across the whole process of innovation is needed.

1.5 **Purpose Statement**

The purpose of this project is to develop and do an initial validation study of an innovation test. The test needs to cover the whole process of innovation and needs to evaluate individual students' abilities at performing each of the tasks outlined by process. This paper will describe the development of the test, including analysis of the content domain, identification of the learning outcomes, item creation, testing of the test, and initial validation.

2 REVIEW OF LITERATURE

This research relies heavily on the previous work of Lewis (2011). In his research, Lewis did an extensive review of the literature on innovation, creativity, and the assessment and measurement of both. He concluded that current innovation and creativity tests were not sufficient to measure students' ability to innovate.

2.1 **Definitions**

Lewis discussed the varying definitions of innovation and creativity as part of his study. He found that there is little consensus on how to define creativity and innovation. This led him to study the processes of innovation as described by innovators. He found that although the processes varied to some degree, they all had some common elements. By looking at the common elements of the innovation process, Lewis bypassed the need to have a specific definition of innovation. In this paper, the definition being used by the Innovation Bootcamp will be used. It is "original and useful ideas implemented successfully." This definition is being used because it is broad enough to encompass the varying definitions of innovation, and puts the focus of the test on the process of innovation rather than specific definitions.

2.2 Current Tests

According to Lewis, current innovation tests primarily measure only part of the innovation process – those that use convergent thinking. The convergent thinking parts of the process focus on analyzing and breaking down ideas or problems. Although this is a very important part of the process of innovation, it is only part of the process.

In contrast, Lewis found that creativity tests tend to focus on the other half of innovation: divergent thinking. Divergent thinking concerns being able to think of many varied ideas based on a given stimulus. This is another part of the innovation process, but again, by itself falls short of measuring the entirety of the innovation process.

This description of divergent and convergent parts of the innovation process is important, though incomplete. The divergence/convergence issue is important, but it is even more important to look at all the parts of the innovation process, regardless of whether they are divergent or convergent. It is not enough to show that a test measures a person's ability to think divergently and convergently; the test must measure a person's ability to carry out all the parts of the process, some of which will be convergent, and some divergent.

Creativity tests often focused directly on divergent thinking (Doolittle, 1990; Houtz & Krug, 1995; Meeker 1985). Other creativity tests measure other aspects of divergent thinking, such as flexibility (Jerome, 1971; Gupta, 1982; Cooper, 1991; Golden, 1975; Torrance, 1999), fluency (Jerome, 1971; Gupta, 1982; Cooper, 1991; Houtz & Krug, 1995, Torrance, 1999), and originality (Jerome, 1971; Cooper, 1991; Houtz & Krug, 1995, Torrance, 1999). Divergent thinking and its specific dimensions are important parts of innovation. They describe how well a person can come up with many varied new ideas. Although the ability to come up with ideas is a

central part of the process of innovation, it does not account for other essential parts of the process.

On the other hand, many of the innovation tests did not measure an individual's ability to perform all of the parts of the innovation process either. Innovation measures that purport to measure the whole innovation process, such as those used in OECD (2005), often looked at the performance of teams, products, or companies. For example, measures in Radosevic and Mickiewicz (2003) evaluated the success of innovation programs in terms of financial outputs, such as sales of a product, or an increase in profits during or after the introduction of an innovation course or program. While these may be useful for management to justify the existence of innovation programs, it does not tell us anything about the improvement in the abilities of the individuals participating in those programs.

2.3 **Innovation Models and Processes**

Because of the need to assess a person's skill at specific parts of the innovation process, it is important to describe the innovation processes and models used by leading innovation educators and consultants. Although the different practitioners use varying language to describe their processes, there were many common elements and similarities across the different processes. These common elements are found in the BYU Innovation Bootcamp model. Because the different groups use similar models and processes, future studies should be done to see if this instrument could be used more generally in innovation education.

2.3.1 The BYU Innovation Bootcamp Model

The five parts of the BYU Innovation Bootcamp model (see Figure 2-1) are: Idea finding, idea shaping, idea defining, idea refining, and idea communicating.

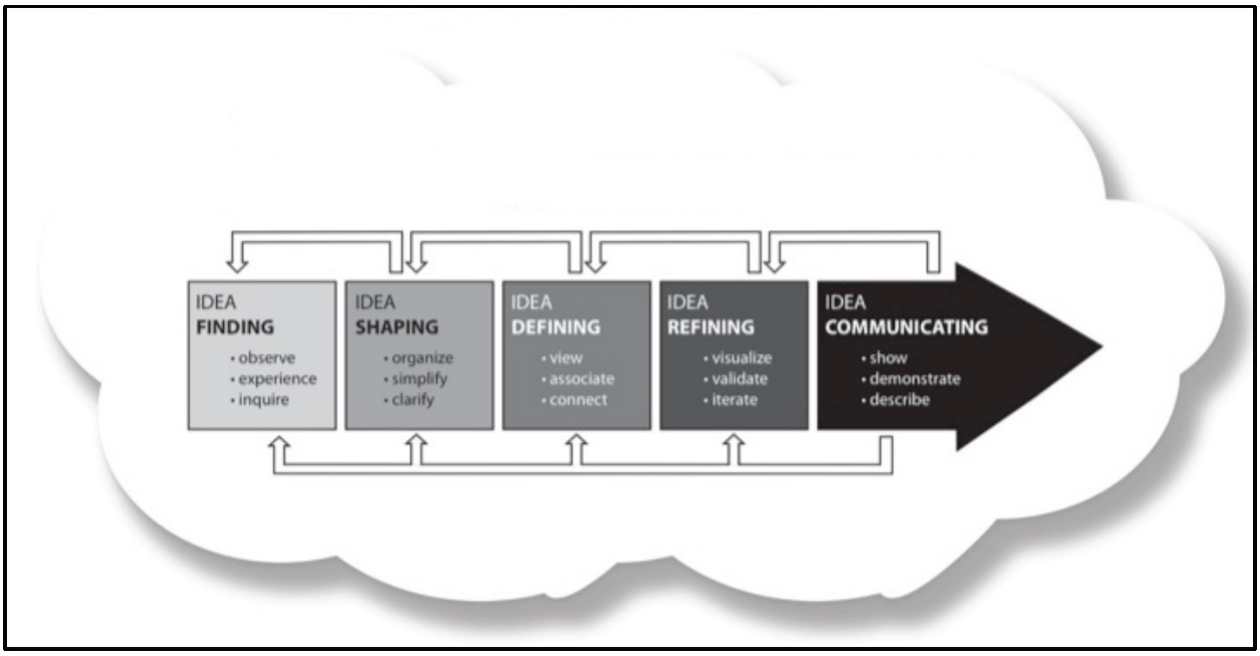

Figure 2-1: BYU Innovation Bootcamp Model

Idea finding involves teaching students to see opportunities for innovation in the world around them. Students in the Bootcamp are taught to take on the role of anthropologist as they observe people. They are taught to actively experience what others are experiencing as to find issues that can be improved upon. Kelley (2005) suggests this approach to innovation and explains how it is used at IDEO. This ties closely to the empathize step in the Stanford D-School innovation process (Stanford, 2010) and also to the behavior of observation described by Dyer et al. (2011). When students actively observe the situations and people around them, they learn to identify opportunities for innovation.

The second part of the BYU Innovation Bootcamp model is idea shaping. In idea shaping, students refine their observations from the idea finding phase. This relates to the define phase of the Stanford D-School model (Stanford, 2010). Stanford describes this as a time to "develop a deep understanding of your users and the design space." The major behavior of this

phase is questioning (Dyer et al., 2011) The goal of this phase is to develop a clear, actionable problem statement. This problem statement guides the rest of the innovation process and gives focus to the participants.

The first 2 parts of the BYU Innovation Bootcamp model comprise what Runco (2006) calls problem-finding. He says that there are multiple problem-finding skills, two of which are "problem discovery" and "problem definition." He cites Getzels' (1975) claim that "the quality of a problem determines the quality of a solution." Problem finding is a critical part of innovation and creates a foundation for the rest of the process.

Idea defining is the third part of the BYU Innovation Bootcamp model. This phase begins the creation of solutions to the problem defined in earlier phases. In this phase, students learn about various methods of ideation and are encouraged to generate a large number of diverse ideas, which Runco (2006) calls fluency. They are taught that they are more likely to have good ideas if they generate many ideas.

Different practitioners use different tools and activities to help ideate. Many of the processes focus on associative thinking (Dyer et al., 2011; Runco, 2006; Mednick, 1962), combining different ideas (often from different fields) to solve the given problem. Dyer et al. (2011) consider associative thinking to be the crux of innovation. They claim that the other behaviors (observation, questioning, idea networking, and experimenting) all lead to creating associations.

After the idea defining phase, the solutions need to be prioritized and refined. Idea refining is the next phase. It is similar to the idea shaping phase, but rather than narrowing down observations to a problem statement, students narrow down solution ideas to a single solution. This is done by choosing the best ideas among those generated in the idea defining phase and by

prototyping and testing them. Experimentation is one of the behaviors identified by Dyer et al. (2011) and fits with the prototype and test steps in the Stanford D-School (2010) process. This phase is important because when many potential solutions are generated in previous steps, the best solutions need to be chosen and refined. By testing the solutions, innovators can see which ones work and how to improve them.

The last phase in the BYU Innovation Bootcamp model is the idea communicating phase. Lewis (2011) found that this is the one part of the process that is unique to the BYU Innovation Bootcamp model. Although this phase is not explicitly mentioned in the other processes studied, it is implied in all of them. All other practitioners of innovation communicate, share, or sell their innovations to others. Rogers (2003) described how innovations diffuse through a community and showed that communication is central to that diffusion. He described how innovations are adopted early by a small part of the population and are diffused to others in the population by various modes of communication.

2.4 **Literature Review Conclusion**

This review describes the lack of adequate innovation assessments to evaluate the innovation skills of individuals. Current tests either do not test subjects on the whole process of innovation, or do not test an individual's skills. This review also describes the process of innovation that the test will cover. It serves as an explanation of the content domain of the assessment, which is necessary to creating a table of specifications and developing the test.

3 METHODOLOGY

This section describes the development of the Innovation Test Instrument (ITI) including identifying learning outcomes of the course, creating a table of specifications for the test, and creating test items. It also describes the testing procedures used to collect initial evidence of form equivalence and validity.

3.1 Identifying Learning Outcomes

The creators of the Innovation Bootcamp curriculum identified five major phases in the process of innovation. These are described in detail in the literature review of this paper. These phases are used to identify the learning outcomes for the Innovation Bootcamp.

The first learning outcome is that students identify opportunities for innovation from a given context. This outcome combines the first two phases of the BYU Innovation Bootcamp model under the umbrella of problem finding. Runco (2006) describes how there are various tools and techniques that fall under problem finding. By focusing on the broader outcomes rather than the particular tools and skills, students can use whatever techniques they want. This outcome focuses on seeing if a student can identify opportunities for innovation, without concern for how they do it.

The second learning outcome is that students create many and varied ideas to solve problems. This outcome tests an individual's fluency, the ability to come up with many and

varied ideas. Fluency has long been an indicator of an individual's level of creativity. By using fluency in a specific context (problem solving, in this case), this outcome targets an individual's ability to create ideas that are useful in the given context.

The third outcome for the course is that students evaluate the ideas based on originality, usefulness, and feasibility. In the Innovation Bootcamp curriculum, innovation is defined as original and useful ideas, implemented successfully. Students should be able to decide whether the ideas they have had fit that definition in order to know which ideas to focus on as they refine and experiment with the ideas.

The fourth outcome is that students can effectively communicate their ideas to others. They should be able to clearly and concisely explain the need for their innovation and the benefit of it. They need to convince readers of the value and impact their innovation will add to those who use it.

3.2 **Table of Specifications**

Based on these learning outcomes a table of specifications was made (see Table 3-1). A table of specifications for creating an assessment is a two-way chart that organizes what the assessment is meant to measure (Miller et al., 2009). It is a common way to visualize what is being measured and shows the proportions of test questions (and test points) that are given to each learning outcome in the course. In these tables the course content areas are arranged along one axis of the table and the cognitive processes are arranged across the other axis (Miller et al., 2009). The cognitive processes used in this instrument's table of specifications are from Bloom's Revised Taxonomy (Anderson et al., 2000).

Table 3-1: Table of Specifications

	Remember	Understand	Apply	Analyze	Evaluate	Create
Opportunity Recognition				2		
Ideation						2
Idea Refining					2	
Communication			2			

The table of specification shows that the test had two items per outcome. Two items per outcome are not necessary to check for equivalent forms, but it was decided that having more equivalent items could make future validation studies easier (after doing initial testing, the number of items per outcome was reduced to one in order to reduce the effects of test fatigue, see below).

3.3 **Item Development**

After creating the table of specifications, items for each outcome were created. This assessment had four item types. The first item type corresponded to the first learning outcome and tested students' ability to identify problems from a photograph. The students were graded on their ability to identify as many problems as possible in the photograph. Answers were given higher weight if they are less common.

The second type of item is similar to the first. It corresponds to the second learning outcome. Students were given a problem statement and asked to generate as many solutions as possible. They were also given more points for answers that come up less frequently than others. This grading scheme is used in other instruments to measure creativity (Torrance, 1969). Torrance uses shapes that students identify and elaborate on and they are awarded for having many answers and unique answers. The difference between the items for this innovation

assessment and the items in Torrance is that these items are focused on problems that people have with their products or environments. So where Torrance items show an abstract shape, these items show an actual problem that could be solved.

The first two items types for this test were designed to be easily changed for future tests. In the first, a subject generates problem statements from a given photograph, and in the second, a subject generates solutions from a given problem statement. It was expected that it would be difficult to achieve equivalent item difficulties for these items on the first attempt. Subjects would likely find it easier or harder to find problems (or solutions) based on the given stimuli. For this reason the items have been designed to be easily modified for future testing. With this item design, photographs (or problem statements) can be easily switched out and tested until equivalently difficult stimuli can be found. In this study, the current items were be tested to see how equivalent they were. Future studies can then easily modify the items to get better equivalence, if needed.

The third type of question tests the students' ability to evaluate ideas. In the innovation process, students come up with many ideas to solve a certain problem. After they have generated those ideas, they have to decide which ideas to pursue and refine. The ability to decide which ideas will be best is what is tested in the third type of items. In this item type students were given a problem statement and four possible solutions. They were asked to rank the solutions according to the definition of innovation used by the Innovation Bootcamp: Original and useful ideas that can be implemented successfully. Their rankings were matched against the rankings that the Brigham Young University Industrial Design faculty made.

In order to create a key for the innovation ranking items, five Industrial Design professors were polled using the items from the assessment, which include the criteria for ranking the

innovations. The key was made by giving points to the innovations that the professors ranked highly. With the totaled scores, an overall ranking could be created that combined the rankings of all the professors (see section 4.2.3 for the professors' responses). Then the students' rankings could be compared to overall rankings when the tests were scored.

The fourth item type tests the students' abilities to communicate their ideas to others. In this item they are asked to create a pitch for the innovation that they ranked first in the previous ranking item. The pitches need to be concise, persuasive, and need to communicate the value of the innovation. In the test, the students are limited to 700 characters in order to maintain conciseness and are graded on persuasiveness and ability to communicate the value of their innovation.

In order to grade this item, two raters were used. Raters followed a provided rubric (see Table 3-2). Raters were trained on how to use the rubric and then graded five questions. They graded pre-selected responses that were considered (by the author) good, mid-grade, and poor in order to ensure that raters could be reliable at different levels of performance. The raters discussed any areas that they disagreed upon. After the first five responses and their discussion, the raters graded five more responses and discussed the scores. This continued until raters achieved a correlation greater than 0.75, because an inter-rater reliability above 0.75 is considered "excellent" (Cicchetti, 1994). After the raters graded all responses, the inter-rater reliability was estimated for all the scores.

Table 3-2: Rubric for Communicate Items

Explain problems: How well does this explain the problem?		
Fails to explain the problem 0	Alludes to the problem 1	Clearly explains the problem 2
Explain solutions: How well does this explain how the solution works or solves the problem?		
Fails to explain the solution 0	Explains, but not clearly 1	Clearly explains the solution 2
Persuasiveness: How well does this convince you of the benefit of the innovation (overall score)?		
This does not convince me 0	This makes me interested 1	This convinces me 2

3.4 **Testing Procedures**

In order to collect initial evidence of validity and form equivalence of the instruments, the test was administered to the students of the Innovation Bootcamp from winter semester 2012. During this semester there were five sections of the Bootcamp with 20 students in each section. As a preliminary check, the first three sections received the test. After they responded, the results were analyzed and revisions were made to the test. The revised test was then given to all 100 students from all sections of the Bootcamp from winter semester. For the full test, students were instructed that the test would be a contest. The students competed for prize money that would be awarded to the students with the highest scores on the test. This was done in order to raise the stakes for the test enough to prompt maximum performance. Then the full test results were analyzed, and suggestions for future studies were made.

In this study, various types of validity evidence were gathered. Content-related evidence was gathered as part of the review of the literature, the comparison of the Innovation Bootcamp with other innovation models, and the description of the alignment between the Bootcamp curriculum and the ITI. Construct-related evidence was addressed in the revisions that were

made between the two rounds of testing, and the description of the methods could be used as initial evidence that could support construct validity once other studies have been performed. Some evidence of face validity was observed through students' enthusiasm for the test and curiosity about the test and how it was graded. Criterion-related evidence was gathered indirectly, with informal observations that connected high test performance to high performance in the Bootcamp. Because the results of this test will have no impact on the students taking it, consequence-related evidence was not a major issue in this study. See section 5.1.1 for discussion of the findings of this study in relation to the different types of validity evidence.

3.5 Revisions to the ITI After Initial Test

After the initial test, the results were analyzed and revisions to the ITI were made in order to improve the test. Full results and analysis of the test will be reported in the results chapter of this paper, but here the revisions will be discussed.

3.5.1 Lack of High Performance

The biggest problem with the initial test was that the subjects did not achieve maximum performance. Few of the subjects finished the test. Others quickly went through the items without giving much thought to them. This likely happened for a couple of reasons. One is test fatigue. Subjects' performance dropped off significantly the longer they spent on the test. This was remedied by making the test shorter. The original length of the test was longer so that there would be a larger item bank for future testing. This proved infeasible for this study because the subjects could not maintain concentration over the large number of items.

The second reason for inadequate performance was that the stakes were not sufficiently high to prompt maximum performance. In order to resolve this issue, the second round of testing

was done as a competition. Prizes were offered to subjects who scored more highly on the tests. The highest-scoring subject would receive $100, the next two highest would receive $50, the next two received $25 and the next ten received vouchers for a free smoothie. This would presumably be enough of an incentive to prompt maximum performance.

Fixing these two problems with the test strengthens evidence of construct validity. Problems with fatigue and lack of incentive hurt the construct validity of a test. They show that problems in the test procedure affected scores enough that they do not accurately describe a person's ability to perform the tasks. By fixing these problems, a stronger claim of construct-related evidence can be made.

3.5.2 Lack of Variation in Responses to Problem-Finding Items

In the first version of the test, photographs were used in the problem-finding items as stimuli for the subjects to find problems with. The photographs were taken of specific problems. It was hoped that students would see the picture and identify many problems in it. For example, subjects were given a picture of a person sleeping on one of the public couches on campus (see Figure 3-1). After looking at the results of these items, we found that there was little variation in the responses. Most subjects mentioned a few obvious problems in the photograph but failed to find anything else. This weakened the item's power to discriminate between people that could find problems well and those who could not. By fixing this problem, evidence of construct validity was strengthened because the item was more able to better target varying levels of the construct.

Figure 3-1: Example of Problem-Finding Photograph

In order to solve this problem, the new version of the test had wider-angle photographs of rooms (see Figure 3-2). This gave the subjects more opportunities to notice a larger number of problems. The hypothesis was that giving the subjects more to look at would allow for a greater variety of answers and give researchers a better idea of the subjects' ability to find problems.

Figure 3-2: Example of Revised Test Photograph

3.5.3 Communicate Items

Another major problem with the initial instrument was in the communication questions. It was evident from many of the answers that the students did not understand what was being asked of them. Many failed to describe the problem or solution well. They had a hard time describing what the problem was that they were trying to solve. They also did not realize that they needed to describe how the solution worked. This may have been because they were given the problem and solution in the previous problem. To fix this, the communicate items were moved to follow the solution-generating items. Rather than trying to pitch a solution that was given to them, the subjects were now asked to pitch their favorite of the solutions they came up with.

Also, the instructions for the communicate items were changed to be more clear. Rather than just tell the students to create a pitch for the idea, they were told what needed to be in their pitch. The new instructions told them to describe the problem, describe the solution, and to be persuasive.

These changes also strengthen construct-related evidence because they show that scores are not affected by other factors than the students' ability to communicate about their innovations.

3.6 Test Form Equivalence

Because a major part of this study is to create equivalent forms that can be used for pre- and post-testing, both forms of the test were given to the students at the same time. To find the forms equivalent, corresponding items should have similar means and standard deviations for the same group of test subjects. Also, student rankings by total score should be the same for both

forms of the test. That is, if a student scores highly compared to others on form A, that student should score highly compared to others on form B.

4 RESULTS

This section describes the data collected from the two rounds of testing and gives explanation of the results. It describes the overall results of each round of testing, the results and analysis of each item type, and the changes that were made after the first round.

4.1 Overall Results of Initial Test

The initial test was given to the first three sections of the Innovation Bootcamp from the winter 2012 semester. Of that group there were 24 subjects who responded to the invitation to participate in the test. Half of the subjects (Group A) first received form 1, and then completed form 2; the other half (Group B) first received form 2, and then form 1. The groups were selected by splitting each of the Bootcamp sessions in half and putting each half in one of the groups. This was done to adjust for any effect of the order of the test forms. Table 4-1 shows the individual scores and the means and standard deviations of the groups.

These data show that scores decreased as students spent more time on the test. In group A, mean scores on the forms decreased from 44.92 to 30.92. In group B, they decreased from 51.83 to 46.33. It is unclear why the decrease in scores was more pronounced in Group A than Group B, but the decrease is large enough to be a concern. This decrease is likely attributed to two factors: 1.) test fatigue, and 2.) lack of incentive.

Table 4-1: Summary of Overall Scores

	Overall totals	Total from 1	Total from 2
1->2 Group	158	78	80
	119	64	55
	109	59	50
	91	53	38
	76	44	32
	72	39	33
	68	41	27
	67	49	18
	52	29	23
	41	29	12
	32	29	3
	25	25	0
Mean	75.83	44.92	30.92
SD	36.95	15.67	21.88
Correlation	.93		

	Overall totals	Total from 1	Total from 2
2->1 group	166	85	81
	162	77	85
	128	62	66
	118	52	66
	114	44	60
	104	57	57
	100	31	60
	91	42	58
	79	37	42
	55	34	20
	54	35	20
	7	0	7
Mean	98.17	46.33	51.83
SD	43.58	21.60	23.60
Correlation	.86		

Observation showed that subjects became fatigued because of the length of the test and the number of items. For example, many of the subjects did not attempt to complete later items on the second form.

Because of this finding, the test was modified into a significantly shorter version. Originally, each form of the test was going to have two items of each type. The limitation of test fatigue required that the second version of the test have one item per item type on each form.

Another limitation of the results is that many of the students failed to achieve maximum performance on the test items because they were not interested enough in completing the test (not enough incentive). Some subjects skipped essay questions or answered them with only a few words, which was problematic because the test was designed to score participants based on subjects' maximum performance of cognitive tasks. In the initial trial of the test, stakes were not

high enough to prompt maximum performance. Consequently, students were given incentives for performance on the second version of the test.

4.2 Analysis of Individual Items

Analysis of the scores and responses of individual items were used to gather evidence of validity and to find ways to improve the items for future tests. Even though the initial test's issues of length and test fatigue limited what could be learned from these results, there were still important things shown. Some of the items did not perform as expected, and were revised for the second round of testing. The problem finding items did not generate a large enough variety of responses and were modified. Also, the communicate items needed better instructions and were modified to help subjects understand better what was expected of them.

4.2.1 Analysis of Problem-Finding Items

In the problem-finding items, subjects tried to identify problems from photographs provided in the test. After the subjects responded to the items, a rater counted all of the responses to find which responses were more common than others. Figure 4-1 through Figure 4-4 show the pictures used in each item and Table 4-2 through Table 4-5 show the corresponding response counts.

Figure 4-1: Photograph from Man on Couch Problem-Finding Item

Table 4-2: Response Counts for Man on Couch Item

Response	Frequency	Score
light in eyes	17	
bad schedule/too tired	13	1
discomfort	10	
cold	8	
no place to sleep	6	2
couch not designed for sleep	5	
unsafe	3	
people shouldn't sleep there	3	
not sleeping well at home	3	
ugly couch	2	3
looks funny	2	
sleeping in public	2	
no friends	1	
hard to breathe	1	
noise	1	4
couch smells	1	

Figure 4-2: Photograph from Leaky Drain Problem-Finding Item

Table 4-3: Response Counts from Leaky Drain Item

Response	Frequency	Score
leaky drain	17	
cords in way	11	1
calcification/corrosion	9	
cords get wet	5	
wet floor hazard	5	2
pipe material	5	
no drain	3	
clogs	3	
lack of maintenance	3	
lack of supplies	3	
can location	3	3
can will fill	3	
rug	3	
can needs liner	2	
seal wears out	2	
design of can	2	
location of outlets	1	
mess	1	
no money to fix	1	4
no time to fix	1	
location of pipes	1	
ugly room	1	

Figure 4-3: Photograph from Printer Problem-Finding Item

Table 4-4: Response Counts from Printer Item

Response	Frequency	Score
printer theft	11	
paper theft	9	1
can't access maintenance	9	
printer too big	5	
make printer more compatible	4	
lack of trust	4	2
unauthorized tinkering	4	
small screen	4	
ugly	4	
paper inaccessible	3	
use of paper	3	
location	3	
not enough maintenance	2	
paper jams	2	3
ink theft	2	
printer moves	2	
hard to reach paper on top	2	
printer tipping over	2	
budget constraints	1	
overheating	1	
people get frustrated	1	
printer can be unplugged	1	
lack of portability	1	4
screen glare	1	
ink runs out	1	
separating print jobs	1	
screen accessible	1	

Figure 4-4: Photograph from Street Cracks Problem-Finding Item

Table 4-5: Response Counts from Street Cracks Item

Response	Frequency	Score
asphalt cracks	11	
ugly	8	1
temperature cracks	6	
tar doesn't work	5	2
weight cracks	4	
unsafe	4	
moisture enters	4	3
not weather proof	3	
bumpy	3	
poor base	2	
infrequent maintenance	1	
unbalanced road	1	
lazy government	1	
budget problems	1	
takes a long time to fix	1	
patch comes off	1	4
unused space	1	
spills on road	1	
tar misaligned with cracks	1	
ground shifts	1	
roots under road	1	
lines are confusing	1	

The mean scores and standard deviations are shown in table 4-10. The table shows the overall means and standard deviations as well as the means and standard deviations of the two test groups.

These statistics show that there is a significant order effect. They show that subjects tended to perform better on items that they completed earlier in the test. This makes establishing equivalence between the items difficult because it is unknown whether the change in scores is a result of them being more difficult or if it is because of the order in which the subjects completed the items. Notwithstanding the order effect, some claims can be made about the difficulty of the items. Both groups scored higher on the printer item than the street cracks item. Because these items were placed in the same section of the test, this shows that they are not likely to be equivalently difficult items. The other scores are inconclusive because even though the man on couch and leaky drain items were in the same section of the test, Group A performed better on the man on couch item, and Group B performed better on the leaky drain item.

Table 4-6: Summary of Statistics for Problem-Finding Items

Overall	Man on couch	Leaky drain	Printer	Street cracks
Mean	7.75	7.88	7.33	6.71
Standard Deviation	3.94	5.24	5.91	5.59
Group A	Man on couch	Leaky drain	Printer	Street cracks
Mean	9.17	8.17	6.58	5.75
Standard Deviation	4.47	6.15	5.68	5.83
Group B	Man on couch	Leaky drain	Printer	Street cracks
Mean	6.33	7.58	11.08	7.33
Standard Deviation	2.66	4.11	6.78	5.47

The response count results show how the subjects responded to the problem-finding items. It shows how divergent the subjects' answers were for these items. The "man on couch"

and “street cracks” items showed less divergence in their responses. The other two items had more items, but it was still unclear if other types of photographs would give subjects a wider range of responses. This led to the decision to test different photographs in the second round of testing. In this initial test, problem-finding photographs were taken of specific problems similar to the ones that students identify in the Innovation Bootcamp. The second version of the problem finding items had pictures that were taken of scenes from a home, without focusing on specific problems. It was hoped that these photographs would give subjects more opportunities to identify a wider range of problems and that having to identify problems from a broader scene would be closer to the experience of problem findings that students face in the Bootcamp and that innovators face in real-world practice.

4.2.2 Analysis of Solution Items

The solution items give subjects problem statements and ask them to generate as many solutions as they can. Scoring of these items follows a similar procedure to the problem-finding items. Students receive points for the solutions they generate, with more points for less common responses. Table 4-7 through Table 4-9 show the problem statements and the corresponding subject responses.

Table 4-7: Response Counts for Garbage Liner Item

Garbage can liners often slip down inside of the cans when they are full of garbage.		
Response	Frequency	Score
strap/band around can	15	
clips on top of can	15	
liner-less cans	14	
bigger bag	14	
create stiff bags	11	1
attach can to chute	11	
elastic bag rim	11	
disposable trash cans	10	
tie knot in liner	9	
sticky rim	9	
don't fill can too much	7	
hooks on lip of can	7	
stretchy bag	7	
incinerators	6	2
sensor to tell when full	5	
automatic/self-sealing bag	4	
velcro rim	4	
drawstring	4	
shorten can	3	
magnetic rim	3	
don't throw things away	3	
layered liner that peels away	3	
Compactor	3	
conical shaped cans	3	
tabs that hold bag	3	
Recycle	3	
slip guard on can	2	3
don't use garbage can	2	
weighted draw string	2	
bags fitted to can	2	
can with hole in bottom	2	
moving support for bag	2	
non-slip bags and cans	2	
lid that pinches bag	2	
hire a maid	2	
continuous bag tube	1	
static cling bag	1	
stack things to support bag	1	
dissolve trash	1	
pigs eat trash	1	
put a max fill line in bag	1	
horizontal can	1	
hole in can that holds excess bag	1	4
robot trash pickup	1	
hang bags without cans	1	
automatic can cleaner	1	
separate can for heavy things	1	
trash teleportation	1	
expanding lid	1	
square cans	1	
washable cans	1	
suction cup rim	1	
vented can	1	

Table 4-8: Response Counts for Headphone Item

Headphone wires get tangled in people's pockets.		
Response	Frequency	Score
retractable spooler	29	
wireless headphones	24	1
wrap guiding device	20	
part of clothing	14	
cover wires with slippery	8	
straight/stiff wires	8	
speakers instead	7	
don't put them in pockets	6	2
special pocket in pants	6	
pocket wire case	5	
stretchable cords	5	
Bluetooth	5	
wrap around ipod	4	
implants in ears	4	
magnetic casing	3	
thicker wires	3	
tie wires in loops	2	3
directed audio speakers	2	
ear buds are the ipod	2	
don't have pockets	2	
flat wires	2	
place to wrap on device	2	
Holster	2	
twist ties	1	
clothespin	1	
clip	1	
teach people how to wrap them	1	
shorten wires	1	
velcro strap to hold wires	1	4
ambient music	1	
get bigger pockets	1	
get other objects out of pockets	1	
make cords tangled	1	
color code wires	1	
public performance instead of ipods	1	
zipper headphone cord	1	
wind wires differently	1	

Table 4-9: Response Counts for Corner Cutting Item

People often cut across the lawn in places around campus which leaves ugly dead patches in the grass.		
Response	Frequency	Score
make more paths	17	
fences	12	
signs/advertising	11	1
plant barrier	10	
wear-resistant grass/turf	9	
raise grass	6	
guards	6	
stepping stones	4	2
get rid of all grass	4	
round corners	4	
ropes	3	
spikes in grass	2	
busses/transportation	2	
make honor code rule	2	
rocks instead of grass	2	
curbs	2	3
guard dogs	2	
move buildings closer	2	
paint dead grass	2	
ziplines/swings	2	
pay a fee	2	
rewards for not crossing	1	
sound wave barrier	1	
research	1	
make crossers work	1	
green gravel	1	
emerald lawns	1	
force field	1	
improve walks	1	4
give people more time	1	
make it muddy	1	
build a fountain	1	
use wood chips	1	
have less students	1	
bridges	1	
security cameras	1	
water borders	1	
make grass stain	1	

Table 4-10: Response Counts for Bakery Item

A local supermarket has to discount their leftover baked goods after they are a day old.		
Response	Frequency	Score
bake less	11	
decrease prices	9	1
better prediction	9	
donate leftovers	8	
bake to order	7	
Advertise	7	
display/sell better	5	2
reuse in other recipes	5	
better preservatives	4	
speed up baking	3	
give out samples	2	3
change laws	2	
increase humidity	1	
don't sell other things	1	
better storage	1	4
burn it	1	
combine with other items	1	
lie	1	

These responses show that some of the items gave the subjects greater opportunities for different answers than others. The bakery item performed particularly poorly in this regard. It did not generate a very large number of different responses from the subjects. The garbage liner item performed best, then the headphone item, and the corner-cutting item, in that order. Other than the bakery item, these items garnered more responses than the problem finding items.

The mean scores and standard deviations of the solution items are shown in Table 4-11. The table shows the overall means and standard deviations as well as the means and standard deviations of the two test groups.

Table 4-11: Summary of Statistics for Solution Items

Overall	Garbage liner	Headphone	Bakery	Corner cutting
Mean	7.33	6.71	5.71	9.88
Standard Deviation	5.91	5.59	4.25	8.91
Group A	Garbage liner	Headphone	Bakery	Corner cutting
Mean	5.50	5.83	4.50	5.33
Standard Deviation	2.25	3.08	3.75	4.17
Group B	Garbage liner	Headphone	Bakery	Corner cutting
Mean	9.17	7.58	6.92	14.42
Standard Deviation	7.61	7.17	4.37	10.00

As with the problem finding items, it is difficult to determine item equivalence based on the data shown here because of the order effect due to test fatigue. These data show that for both groups the bakery item was the most difficult. The other scores do not conclusively describe the equivalence of the other items.

The data from the solution items show that they performed better than the problem-finding items. In most of the items, the subjects gave a larger number of different responses than in the problem-finding items. Because of this, two of these items were chosen for the second round of testing. The garbage liner and headphone items were chosen for more testing because their means were closer than the others and they had a large number of different responses.

4.2.3 Analysis of Ranking Items

The ranking items gave subjects a problem statement and four potential solutions. Subjects ranked solutions using the Innovation Bootcamp's definition of innovation: original and useful ideas implemented successfully. Before the test was administered, the ranking items were given to four Industrial Design faculty. Their rankings were used to create a key to grade the students' scores by summing the point values from their rankings and then ranking the totals.

Table 4-12, Table 4-14, Table 4-16 and Table 4-18 show the final rankings based on the professors' responses with the problem statements and solutions and Table 4-13, Table 4-15, Table 4-17, and Table 4-19 show how each of the professors ranked each item. Table 4-20 shows the overall and group means and standard deviations for the ranking items.

Table 4-12: Problem Statement and Experts' Rank Order for Bike Seat Item

Bike seats are often exposed to the weather and become wet or absorb water, which causes discomfort to the rider.	
1	A plastic cover with elastic around the edge (like a hairnet) that protects the seat from becoming wet.
2	The seat has ridges that channel water away from the rider and off the surface of the seats.
3	Small, removable seat that the rider can take with them while not riding the bike.
4	A wide fender that folds down to protect the rider from water that splashes from the tire while riding. While not riding, the fender folds up and shields/cover the seat from the weather.

Table 4-13: Expert Responses for Bike Seat Item

	Plastic cover	Fender	Ridges	Removable
Professor 1	1	3	4	2
Professor 2	1	4	3	2
Professor 3	3	4	1	2
Professor 4	2	3	1	4
Total	7	14	9	10
Rankings	1	4	2	3

Table 4-14: Problem Statement and Experts' Rank Order for Toilet Item

People don't like to sit on public toilets. How do we make them more sanitary?	
1	A toilet that automatically sprays disinfectant after every flush.
2	Seats with multi-layered tissue, one layer is removed after each use.
3	Toilet with no seat and people hold on to handrails and squat down.
4	Removable toilet seats with a seat washer in the bathroom.

Table 4-15: Expert Responses for Toilet Item

	Spray	Removable	Tissue	No Seat
Professor 1	2	4	1	3
Professor 2	2	3	4	1
Professor 3	1	4	2	3
Professor 4	2	3	1	4
Total	7	14	8	11
Rankings	1	4	2	3

Table 4-16: Problem Statement and Experts' Rank Order for Lawnmower Item

When people mow their lawns, the grass clippings take up a lot of space after they are done.	
1	A tank people put in their yard that chemically breaks down the grass into fertilizer.
2	A service that collects clippings and converts them to fertilizer.
3	A grass that doesn't grow longer than the desired length.
4	A lawnmower that burns the grass clippings after they are cut.

Table 4-17: Expert Responses for Lawnmower Item

	Burns	Non-growing Grass	Collection Service	Fertilizer Tank
Professor 1	4	3	1	2
Professor 2	4	3	2	1
Professor 3	4	1	3	2
Professor 4	3	4	2	1
Total	15	11	8	6
Rankings	4	3	2	1

Table 4-18: Problem Statement and Experts' Rank Order for Outlet Item

In classrooms, people often need power in places where there aren't power outlets.	
1	Retractable extension cords built in to the walls.
2	Outlets on a track system so that you can move the outlets around the room.
3	Have more charging stations so people don't need outlets.
4	Wireless power supply.

Table 4-19: Expert Responses for Outlet Item

	Retractable	Wireless	Track System	More Stations
Professor 1	1	4	2	3
Professor 2	1	4	2	3
Professor 3	1	2	3	4
Professor 4	1	4	2	3
Total	4	14	9	13
Rankings	1	4	2	3

Table 4-20: Summary of Statistics for Ranking Items

Overall	Bike Seats	Toilets	Lawnmowers	Outlets
Mean	4.92	6.71	3.92	2.88
Standard Deviation	3.08	2.78	2.83	2.11
Group A	Bike Seats	Toilets	Lawnmowers	Outlets
Mean	5.58	6.42	3.67	3.00
Standard Deviation	2.98	2.87	3.27	2.24
Group B	Bike Seats	Toilets	Lawnmowers	Outlets
Mean	4.25	7.00	4.17	2.75
Standard Deviation	3.03	2.65	2.27	1.96

The data show that the outlet item is more difficult than the other items because both groups did significantly worse on it than on the other three items. The lawnmower item also appears to have scored much lower, but in group B, the lawnmower item scored close to the bike seat. Group A and the overall scores show the lawnmower item lower. Because of this, the bike seat and toilet items were chosen to be retested in the second test. By testing these two items more, it could be better established whether or not they are equivalent. Testing them again in a shorter test could also lessen the effect of test fatigue and give researchers a clearer view of the equivalence of these items.

4.2.4 Analysis of Communicate Items

The communicate items followed the ranking items in the assessment. The communicate items asked the subjects to create a pitch for the innovation that they ranked highest on the second ranking item. They were asked to create a convincing pitch that would persuade others to adopt the innovation they chose. Table 4-21 shows the overall and group statistics for the communicate items from each form of the instrument.

Table 4-21: Summary of Statistics for Communicate Items

Overall	Form 1 Item	Form 2 Item
Mean	4.33	3.63
Standard Deviation	3.57	3.84
Group A	Form 1 Item	Form 2 Item
Mean	4.25	2.08
Standard Deviation	3.42	3.28
Group B	Form 1 Item	Form 2 Item
Mean	4.42	5.17
Standard Deviation	3.71	3.74

These data show that subjects in both groups performed poorly on both of the items. The total points possible on the items were 12, and the means of the responses were less than half of that. A few problems with the items were observed upon looking at individual responses.

The first problem was that many of the subjects gave very limited responses to these items. It appeared that subjects did not care enough about the test to go through the effort of constructing a good response to this item. Many did not finish the item. Researchers attempted to remedy this problem in the second round of testing by making the second round a competition with prizes to the subjects that scored most highly on the test. By giving the subjects more of an incentive to perform, researchers hoped to prompt better responses, especially on the communicate items (see section 4.3.3 for the results of the second round of testing communicate items).

The second problem was that most subjects wrote the pitch as if the rater already understood the problem statement and the solutions. It was difficult for them to write about the problem and how the innovation fixed it when they were given both the problem and the solution. For this reason, in the second version of the test, communicate questions were tied to the solution questions rather than the ranking questions. After the students generated their

solutions from the given problem statement, the communication item was placed next so that students could explain the benefits of the innovation that they came up with rather than the innovation they were given.

The third problem was that subjects did not always understand what they were supposed to write in the pitch. Some subjects described their rationale for choosing one of the responses over the others. Others failed to mention what the problem was or how their choice would solve that problem. It seemed that the subjects did not understand what was expected of them on these items. To remedy this problem, clearer instructions were created for this item. The first version of the test had these instructions on the communicate items: "For the idea that you picked as best in the previous question, write a pitch to convince people of the benefits of the solution." The new instructions read: "Choose your favorite idea from the previous question and write a pitch about it. Include the following: 1) Describe the problem you are trying to solve. 2) Describe how your solution fixes the problem. 3) Be convincing. Persuade people that your solution is a good one."

One aspect of these items that worked well was their rating. Using the grading rubrics, the raters scored the items with high reliability levels: 0.94 for the item from form 1 and 0.97 from form 2. Cicchetti (1994) said that reliability scores above 0.80 are considered "nearly perfect." This could be due to the training procedure explained in the methods chapter of this paper, but is also a result of so many of the responses being poor (raters easily agreed on responses that were severely lacking). Testing in the second version (when the stakes were higher) was expected to give lower inter-rater reliability scores because there would be fewer very poor responses.

4.3 **Overall Results of the Second Test**

With the new revisions made based on the analysis of the initial test, the second version of the test could be administered to the group. The second test had nearly half as many items as the original test to limit test fatigue. The test was offered to all 100 students of the Innovation Bootcamp from the winter 2012 semester. They were promised prizes for the top 15 responses, with the top score receiving $100. Having a shorter test and raising the stakes for performance had a drastic impact on the responses. Of the students who were given the opportunity to take the test, 39 responded. All students who responded completed all items on the test and many of the students spent more time on the second version of the test than on the first, even though the second test was half as long. The results of the second round of testing are shown in Table 4-22.

Table 4-22: Summary of Second Test Scores

	Overall totals	Total from 1	Total from 2
Group C	116	62	54
Form 1->2	105	47	58
	84	39	45
	79	42	37
	79	50	29
	75	34	41
	73	38	35
	71	37	34
	71	34	37
	71	38	33
	70	33	37
	66	28	38
	64	33	31
	61	36	25
	59	32	27
	59	29	30
	54	24	30
	50	21	29
	41	22	19
	41	24	17
mean	69.45	35.15	34.30
st dev	17.95	9.74	9.81
correlation	.69		

	Overall totals	Total from 1	Total from 2
Group D	142	61	81
Form 2->1	95	54	41
	92	38	54
	89	42	47
	88	44	44
	83	42	41
	82	39	43
	72	33	39
	70	29	41
	69	27	42
	64	29	35
	63	24	39
	61	28	33
	60	35	25
	58	36	22
	56	26	30
	56	31	25
	53	27	26
	48	25	23
mean	73.74	35.26	38.47
st dev	21.28	9.76	13.28
correlation	.70		

These data show that the order effect was greatly reduced from the initial test. The increased consistency of the scores made the comparisons between the items in the new test more meaningful than in the initial test.

4.3.1 Results for Problem-Finding Items (Second Test)

The problem-finding items on the second version of the test used the same format as the first, but with different pictures. The pictures used in the second version of the test are shown in Figure 4-5 and Figure 4-6. The response counts are shown in Table 4-23 and Table 4-24.

Figure 4-5: Photograph from Garage Problem-Finding Item

Table 4-23: Response Counts for Garage Item

Response	Frequency	Score
Organization of bikes	31	
general storage/org	22	
parking arrangements	16	1
items inaccessible	12	
shelving	10	
lack of space	10	
poor lighting	10	
oil stains/dirty floor	10	2
hooks from ceiling	10	
too many bikes	9	
small door	7	
dirty cars	5	
entrance procedure	5	
see box contents	5	
organize items on shelf	4	
store boots/shoes	4	
snowboard on 1 hook	4	
store unused things	4	
car top carrier on cabs	3	
containers on ground	3	
water pipes	3	3
2 garage doors	3	
location of door	3	
too many boxes	3	
lockers don't shut	3	
basketball hoop	2	
better kind of cooler	2	
bikes scratch cars	2	
trailer in driveway	2	
use of vertical space	2	
convertible top fix	2	
floor seal	2	
bike sizing	1	
cabinet doors open	1	
camera lens	1	
car paint fades	1	
door left open	1	
items could fall	1	
messy driveway	1	
number of hobbies	1	
organize tools	1	4
prioritize projects	1	
promote organization	1	
shape of driveway difficult	1	
space for toys	1	
take many bikes on a trip	1	
VWs break down	1	
yellowing parts on fridge	1	
bike maintenance	1	

Figure 4-6: Photograph for Bedroom Problem-Finding Item

Table 4-24: Response Counts for Bedroom Item

Response	Frequency	Score
bed undone	31	
bookshelf full	29	
clothes on chair	22	
sun through window/blinds	20	
shoes on floor	18	1
poor lighting	18	
One leg on chair	13	
messy table	12	
no room for rackets	11	
ball storage	9	
nightstand full	8	
humidifier	8	
general org/storage	8	2
guitar	7	
basket for cables/games	5	
pillow on floor	5	
towel on humidifier	5	
empty floor space	3	
bed storage	3	
no wall space	2	3
trash can liner	2	
lack of power supply	2	
vertical space unused	2	
heated blanket	1	
cd storage	1	
trash bin location	1	
paper storage	1	
workspace needed	1	
cleaning is no fun	1	
air vent location	1	4
sore throat/cough	1	
room temperature	1	
need to show achievements	1	
vacuum under furniture	1	
paint fading	1	
photo lens effect	1	
cup on nightstand	1	

The response counts show that the new problem finding items garnered a much larger variation in the responses. The subjects gave many more and varied responses to the items than they did in the initial test.

The mean scores and standard deviations of the problem-finding items are shown in Table 4-25. The table shows the overall means and standard deviations as well as the means and standard deviations of the two test groups.

Table 4-25: Summary of Statistics for Problem-Finding Items

Overall	Garage	Bedroom
Mean	13.00	9.69
Standard Deviation	6.14	5.89
Group C	Garage	Bedroom
Mean	12.95	9.20
Standard Deviation	4.98	4.12
Group D	Garage	Bedroom
Mean	13.05	10.21
Standard Deviation	7.15	7.27
Item Correlation	0.68	

These data show that the new version of the test had a smaller order effect than the initial test. With the reduced order effect, the equivalence of the items could be studied. The difference between the means of the two items suggests that they cannot be considered equivalent. There appears to be more problems to find in the garage item than in the bedroom item. Looking at the two photographs, there does appear to be more opportunities for innovation in the garage picture because there are more objects in that picture. In order to create more equivalent items, more pictures should be tested and analyzed.

Having more equivalent items could also improve the item correlation. The author hypothesized that the same thing that is causing the difference in means could be negatively affecting the item correlation. It is possible that some people may do better (in relation to the rest of the group) on items with more options (like the garage item) and others may do better with fewer options (as in the bedroom item). Further testing with different prompts will help researchers understand whether the difference in item difficulty affects item correlation. If it is found that difficulty does affect correlation, it may mean that there are multiple factors being measured in these items.

4.3.2 Results for Solution Items (Second Test)

The solution items on the second test were chosen from the original test without changing them. They appeared to be working well in the first test, but it was unclear how equivalent they were because of the order effect, so they were tested again in the second test.

The mean scores and standard deviations of the solution items are shown in Table 4-26. The table shows the overall means and standard deviations and the means and standard deviations of the two test groups.

Table 4-26: Summary of Statistics for Solution Items

Overall	Headphones	Garbage Liner
Mean	8.95	11.15
Standard Deviation	4.85	6.24
Group C	Headphones	Garbage Liner
Mean	8.95	9.60
Standard Deviation	5.04	5.67
Group D	Headphones	Garbage Liner
Mean	8.95	12.79
Standard Deviation	4.64	6.39
Item Correlation	0.46	

The data in this table show that the order effect was also remedied in the solution items. Much like the problem-finding items, the second round of testing gave a clearer view of the equivalence of the items. It showed that the headphone and garbage liner items are not likely equivalent because of the large difference in the means. This data also shows that there was a large difference in performance between the two groups on the garbage liner item. This may be due to the sample size of the groups. Future testing with more items and larger samples should be done to create and identify equivalent items.

As with the problem-finding items, the item correlation may be improved with more equivalent items. It could also be that there are other confounding factors working in these measurements. For example, if a person's past experience had led them to deal with one of these problems before, they may already have solutions in mind for these problems. Future researchers may need to look for problems to use as prompts that are either universally familiar (or universally unfamiliar) to the population that is being tested.

4.3.3 Results for Communicate Items (Second Test)

In the second test, the communicate items were changed to go with the solution items rather than the ranking items. The instructions were also changed to be clearer and describe what the raters were looking for in the items. The overall means and standard deviations and those of the individual groups are shown in Table 4-27.

These data show that even though the communicate items use the same wording, they are not necessarily equivalent. The difference between the scores was more pronounced in group C than in group D. It is not clear why this happened, but it could be that a larger data set is needed to stabilize the results. There may be some statistical anomaly in one of the groups that would disappear with a larger test sample. Some of the differences may come from the differences in

the problem statements from the solution items. More testing would need to be done with different prompts in the solution items. It may be found that solution items with more equivalence could lead to communicate items with more equivalence also.

Table 4-27: Summary of Statistics for Communicate Items

Overall	Headphone pitch	Garbage Liner pitch
Mean	8.62	8.28
Standard Deviation	1.41	1.28
Group C	Headphone pitch	Garbage Liner pitch
Mean	9.10	8.20
Standard Deviation	1.37	1.50
Group D	Headphone pitch	Garbage Liner pitch
Mean	8.11	8.37
Standard Deviation	1.25	0.98
Item Correlation	0.43	

The item correlation may also be improved by making the items more equivalent. Having items that have more similar difficulty may mean that students perform more reliably across the items. Because the communicate items rely so heavily on the solution items, the lack of correlation in the solution items is likely contributing to the lack of correlation in the communicate items. In future studies, researchers should see how the item correlations of the communicate items change as the item correlations of the solution items improve.

Inter-rater reliability for the second test was also high. The correlation between the raters' scores on the two items were 0.76 and 0.74 respectively. This is enough to confidently claim good inter-rater reliability (Cicchetti, 1994).

4.3.4 Results for Ranking Items (Second Test)

The ranking items were chosen from among the items in the first round of testing. The bike seat and toilet items were chosen for the second test. These two were chosen because they were the higher scoring items from the previous test. With the other items scoring so low, there was concern that the items could lose power to differentiate because so many of the scores were low. They were tested in the second test to see how equivalent they are without the order effect and fatigue problems. The summary statistics of the second test ranking problems are shown in Table 4-28.

Table 4-28:Summary of Statistics for Ranking Items

Overall	Bike Seat	Toilet
Mean	4.64	7.21
Standard Deviation	2.90	2.40
Group C	Bike Seat	Toilet
Mean	4.15	7.30
Standard Deviation	2.85	2.22
Group D	Bike Seat	Toilet
Mean	5.16	7.11
Standard Deviation	2.85	2.57
Item Correlation	0.09	

The data in the table show that the order effect and fatigue problems have been resolved, but that the difference in the item difficulties became more pronounced. Both groups performed better on the toilet item than on the bike seat item. More items should be created and tested to find items that are more equivalent.

The item correlation for these items is very low. This shows that there is a serious problem with these items. This problem likely stems from the lack of agreement between expert

rankings (see Sections 4.2.3 and 5.2). With more consensus in the expert rankings, the item correlations will improve because there will be a stronger standard against which students can be compared. As consensus on the correct ranking improves, the items will more consistently discriminate between students who can rank the innovations well and those who cannot.

5 CONCLUSION

This section provides a discussion based on the findings presented in chapter four. It describes (a) the reliability and validity of the ITI – based on student test scores, (b) limitations of the data, and (c) recommendations for future development and validation of the ITI.

5.1 Summary and Interpretation of Findings

Lewis (2011) described the need for an innovation assessment that measures an individual's ability to perform all of the different parts of the process of innovation. A test such as this has value for both industry and academia. This study describes the development and initial validation of such a test. The test is called the Innovation Test Instrument. It is an instrument that measures subjects' skills at performing each part of the process of innovation.

In conjunction with the development of this test, an initial validation was performed. It is not a full and conclusive validation of the instrument, but serves as a foundation for further, in-depth validation studies. In this initial validation study, researchers looked for any major problems with the test and ensured that the test is aligned with the content domain. They also checked for reliability among the raters of the test and for equivalence between the two forms of the test.

Data from the two rounds of testing performed in this study show that the test has great potential for validity in measuring subjects' ability to innovate. Evidence gathered from this

study allowed researchers to improve the test and make a case for initial validity. The ITI appears to measure the subjects' ability to perform the process of innovation. Reliability of the scores on the rater-scored items was high. These findings show that a more in-depth validation study of this instrument would be valuable. This section will discuss the validity and reliability in more detail.

5.1.1 Validity

This study was an early study of validity of the ITI. Although more testing should be done to further establish validity of the scores from this instrument, this study showed that there is a good case for some types of validity-related evidence. This section discusses content-related evidence, consequence-related evidence, construct-related evidence, face validity evidence, and criterion-related evidence of validity.

The description of the processes of innovation in the review of the literature shows the content-related evidence, the degree to which an instrument covers the content within a specific domain (Babbie, 1990). The review showed that the test was aligned with the processes of innovation of leading innovators and educators. The method section also helped to establish a link between the instrument and the content that is to be tested. The review of literature showed that the BYU Innovation Bootcamp curriculum is aligned with other innovation processes and models, and the methods section shows that the ITI is aligned well with the Bootcamp curriculum. Showing the links between instrument, course, and content is evidence of content-related evidence.

Consequence validity describes the "consideration of the consequences of use and interpretation of assessment results" (Miller, et.al, 2009). In this study, the stakes of the test results are very low. Results will not be used to establish grades for students or determine

whether they should be admitted to certain programs or positions. The only real consequence of the results of this instrument in its current form is that results could affect how the Innovation Bootcamp is taught in the future. The results of this instrument should not be used for other considerations without further study.

The methods section also described the development of the test items. It showed that the test items were developed using generally accepted test development practices. This can be a positive initial step in establishing construct-related evidence of validity. Construct validity refers to how well the measurements taken in an assessment relate to each other according to theoretical constructs (Babbie, 1990). Showing that the right methods were used does not establish construct validity on its own, but it does show that construct validity is more likely than if they had not been used. Further effort to establish construct validity should be done and will be described in the recommendations section of this chapter.

Construct-related evidence was also addressed in the revisions that were made between the two rounds of testing. Changing the pictures in the problem-finding items, moving the communicate items, revising the communicate items' instructions, shortening the instrument, and adding incentives were all ways that the author reduce construct-irrelevant variance. That is, they were changed in order to make sure that the results of the test reflect the students' ability to do the tasks and not their ability to do other things such as overcome fatigue or understand unclear instructions. The changes that were made add to the construct-related evidence.

Face validity is a type of validity that refers to how much the respondents perceive that the test is relevant or important (Miller, et.al, 2009). Generally, if test subjects fail to see the importance or relevance of the test, they will be unlikely to participate. The first round of testing showed that the instrument had some face validity for the students of the Innovation Bootcamp.

Even though test fatigue caused results that made some interpretations difficult, the fact that so many students participated as much as they did demonstrates a level of face validity. This improved more in the second round of testing because students were more invested in completing the test well. Some students commented that they enjoyed taking the test or thought it was an interesting way to practice what they had learned in the Innovation Bootcamp. The fact that students felt that the test was relevant to what they had learned is a strong piece of evidence in favor of face validity.

Criterion-related evidence, which is sometimes called predictive validity, refers to how well a measured variable can predict other variables. In this test, a claim of criterion validity would say that scores on this test are a good predictor of how likely a person is to actually be a strong innovator. This type of validity was not studied in detail in this research. Although criterion validity was not formally studied, the author of this study made anecdotal observations that support criterion validity. The author of this research also assists in the instruction at the Innovation Bootcamp. The author noted that the top scorers on the test were also students who had many innovative ideas at the Bootcamp. This, alone, is not enough to establish criterion validity. More research that could support criterion validity will be discussed in the recommendations section of this chapter.

5.1.2 Reliability

In this study, two types of reliability were studied: test form equivalence and inter-rater reliability. Chapter 4 discussed in detail the equivalence of the items. Because of the differences in the means scores of the items, all of the item types in this instrument need additional work before they can be used for pre-post-testing of the Innovation Bootcamp. The design of the items makes future testing of new items a simple process. The photographs, problem statements,

and solutions can easily be changed. Future researchers will need to test different prompts for the items and find prompts that garner more similar means and standard deviations of the scores. Even though this instrument did not achieve form equivalence, it is a strong first attempt that will facilitate future instrument development in the area of innovation assessment.

Although the means and standard deviations of the items show that these items are not equivalent, they can still be used as pre- and post-test items to measure the impact of the Innovation Bootcamp. This can be done by using the data from this sample to compute z-scores for the responses to each item. Z-scores are used because they take into account the different means and standard deviations and allow researchers to compare scores from different samples.

For example, in this study, the garage item had a mean of 13.00 and a standard deviation of 6.14 and the bedroom item had a mean of 9.69 and a standard deviation of 5.89. If a student did the garage item in a pre-test and scored 11, the z-score (in relation to the sample group from this study) would be -0.33. If the student did the bedroom item as part of a post-test, and scored 10, the z-score would be +0.05. In this case, the positive change in the z-score would show that the student performed better on the post-test item than on the pre-test item.

The inter-rater reliability for the communicate items was also tested. In the first round of testing, inter-rater reliability levels were 0.94 and 0.97, and on the second round, 0.76 and 0.74. According to Cicchetti, (1994) inter-rater reliability over 0.74 is considered good. This leads the author to be confident in inter-rater reliability for the scores of the communicate items.

Cronbach's alpha coefficient is often used as a measure of reliability in assessments such as this one. It was not used in this study because the assumptions that Cronbach's alpha makes could not be confirmed in this research. One of the assumptions of Cronbach's alpha is that there is tau-equivalence among the items loading on a factor (Graham, 2006). In order to claim

tau-equivalence, a confirmatory factor analysis is needed. Confirmatory factor analysis was not performed in this study because it requires more samples than were available. In future studies, a confirmatory factor analysis of the results of this instrument would be an important step toward measuring reliability and would have other benefits as described in section 5.3.

5.2 **Limitations of Findings**

After analyzing the data from chapter 4, a few limitations were noted. These limitations should be addressed in future study and validation of the Innovation Test Instrument. One limitation was the sample size for the tests. Some of the response data from the items show significant differences between the groups that cannot be attributed to order effect. These differences may be the result of samples being too small. With large enough samples, the anomalies noted in the data will likely be resolved.

One other limitation was noted in the ranking items. In order to grade the ranking items, they were given to five industrial design professors. These professors ranked the innovations and their rankings were combined to create an overall ranking against which subject responses would be scored. The problem with this is that the professors were not all in agreement on their rankings. This likely caused the low correlation between students' scores on the ranking items. The validity of the ranking items could be greatly strengthened by developing responses that all the experts could agree upon rather than just combining their scores.

5.3 **Recommendations for Future Study**

Based on the findings of this research, there is potential for future studies that can further develop the ITI innovation assessment. Some of these recommendations apply to individual

items from the instrument. Others apply to future validation studies that would be performed on the test as a whole.

The items on the second version of the test had varying levels of equivalence. These items should continue to be modified over time to improve pre-post-testing of the Innovation Bootcamp. The problem-finding items work better when the photographs are of rooms or scenes rather than of individual problems because they gave subjects a wider variety of possible responses. Further study should be done to find scenes that will prompt more equivalent responses so that researchers can compare raw scores instead of z-scores for these items (which will greatly simplify the grading of the tests). Similarly, the prompts in the solution items and the ranking items need more testing with various prompts until prompts that get similar scores can be found.

While researchers continue to improve the equivalency of these items, they can use the current item z-scores to measure the impact of the Innovation Bootcamp (as described in section 5.1.2). The author suggests that future researchers introduce new items alongside these items. That way, they can find other possible candidates for equivalent item prompts while gathering data about the impact of the Bootcamp.

The limitation of the ranking items that was discussed in the previous section needs to be addressed before the test can be used to evaluate the Bootcamp. Demonstrating better consensus among the expert rankings would add to the evidence of validity of these items. This could be done in one of three ways. One would be to get the experts together and have them discuss their rationale for choosing each ranking and then have them come to an agreement about how the innovations should be ranked. The second option would be to continue adjusting and testing the items until the faculty all agree on a ranking. The third option would be to get a much larger

sample of experts and then total all the scores to create the rankings (as was done with the small sample in this study).

Future validation studies should be done to strengthen the claims of validity for this instrument. In this study, construct validity was only studied at a surface level. Confirmatory factor analysis would help establish that the theoretical construct that this instrument attempts to measure are valid. It determines whether or not the factors the test is intended to measure really work the way researchers hypothesize that they do. In this study, four major factors are hypothesized to measure a person's skill at innovation. A confirmatory factor analysis could tell researchers if there are other factors that these items are measuring and if their hypothesized model is right. This type of analysis was not done in this study because it requires a larger data sample than was available. Future studies with larger data sets would allow a confirmatory factor analysis to be done.

Criterion validity is another type of validity that should be studied for this test. This could be done in a number of ways. One would be to use this instrument to test students of the Innovation Bootcamp and then have raters score the performance of the same students as they participate in the course. By comparing the results, researchers could see how well the assessment predicts student performance in the Bootcamp. Studies could also be done that compare students' scores on this instrument with other validated instruments that measure parts of what this assessment does. Scores on this instrument could be compared with scores on other instruments like the ones mentioned in Lewis (2011). Another study would be a longitudinal study of students who take the assessment to see how well it predicts how innovative they are in their later careers. This could be another way of seeing how well the assessment predicts future innovation skill.

5.4 **Conclusion**

This paper described the need for an innovation test to evaluate the effectiveness of innovation courses. It described the content that needed to be tested for and the procedures that the author went through to create the Innovation Test Instrument. It also showed the results of initial validation testing for the test Innovation Test Instrument.

This study is an important step in creating methods of testing students' innovation skills. Based on the testing performed in this study, the Innovation Test Instrument will help researchers understand the effectiveness of the Innovation Bootcamp at improving students' innovation skills. Future testing and development should be done to improve the item equivalency. Even with the items that are not currently equivalent, much of this instrument could be used to begin evaluating the impact of the Bootcamp. By using z-scores for the test items, researchers can compare the scores on the items to see how students have improved as a result of the Bootcamp. Once the problem of the experts' lack of consensus on ranking items is fixed, this instrument will be ready for use.

Overall, there are encouraging signs that testing students' skills at performing specific parts of the innovation process has value in measuring their overall innovation skill. This study can be used as a springboard to more research in the process-based approach to innovation measurement.

REFERENCES

Anderson, L.W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., Wittrock, M. C. (2000). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*, Allyn & Bacon.

Babbie, E. R. (1990). *Survey Research Methods, Second Edition*, Wadsworth Publishing.

Christensen, C.M. (1997). *The Innovator's Dilemma: Why Great Companies fail*, Harper Business.

Christensen, C.M., Eyring, H. J. (2011). *The Innovative University: Changing the DNA of Higher Education from the Inside Out*, John Wiley and Sons.

Cicchetti, D. V. (1994). "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology." *Psychological Assessment*, 6(4), 284-290.

Drucker, P. F. (1985). *Innovation and Entrepreneurship*, Harper Collins.

Dyer, J., Gregersen, H., Christensen, C. M. (2011). *The Innovator's DNA: Mastering the Five Skills of Disruptive Innovators*. Harvard Business Review Press.

Fagerberg, J. (1999). "The Need for Innovation-Based Growth in Europe." Challenge, 42(5), 63-79.

Friedman, T. L., Mandelbaum, M. (2011). *That Used to Be Us: How America Fell Behind in the World It Invented and How We Can Come Back*, Farrar, Straus and Giroux.

Getzels, J. W. (1975). "Problem-Finding and the Inventiveness of Solutions." *The Journal of Creative Behavior*, 9(1), 12-18.

Graham, J. M. (2006). "Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability, What They Are and How to Use Them." *Educational and Psychological Measurement*, 66(6), 930-944.

Howell, B., Wright, G., Fry, R., & Skaggs, P. (2011). "The Innovation Lab." *Proceedings of the International Conference on Engineering Design (ICED11)*.

Hurt, H., Joseph, K., Cook, C. (1977). "Scales for the Measurement of Innovativeness." *Human Communication Research*, 4(1), 58-65.

IDEO (2011). "About IDEO." http://www.ideo.com/about/ (accessed 22 April 2011).

Innosight (2011). "Our Approach." http://www.innosight.com/our_approach/create_or_reshape_process.html?gclid=CIPCytTUsKgCFSUZQgod0BZJHA (accessed 22 April 2011).

Kelly, T. (2005). *The Ten Faces of Innovation: IDEO's Strategies for Defeating the Devil's Advocate and Driving Creativity Throughout Your Organization*. Currency/Doubleday.

Lewis, T. (2011). "Creativity and Innovation: A Comparative Analysis of Assessment Measures for the Domains Of Technology, Engineering, and Business." Master's Thesis, Brigham Young University.

Mednick, S. A. (1962). "The associative basis of the creative process." *Psychological Review*, 69, 220-232.

Miller, M. D., Linn, R. L., Gronlund, N. E. (2009). *Measurement and Assessment in Teaching, 10/E*, Pearson.

Obama, B. (2011). "Winning the Future." *State of the Union Address*. Washington, D.C.

OECD, (2005). "The Measurement of Scientific and Technical Activities: Proposed Guidelines for Collecting and Interpreting Technological Innovation Data." *Oslo Manual*. Paris, France

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*. Free Press.

Runco, M. A. (2006). *Creativity*, Academic Press.

Stanford dSchool (2010). "Bootcamp Bootleg." http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf (accessed 14 May 2012).

Stanford dSchool (2011). "Design Thinking Boot Camp: From Insights to Innovation." http://www.gsb.stanford.edu/exed/dtbc/ (accessed 22 April 2011).

Torrance, E. P. (1969). *Creativity*. Dimensions.

Wagner, T. (2010). *The Global Achievement Gap: Why Even Our Best Schools Don't Teach the New Survival Skills Our Children Need--and What We Can Do About It*, Basic Books.

Wagner, T. (2012). *Creating Innovators: The Making of Young People Who Will Change the World*, Scribner.

Wright, G., West, R. (2010). “Using Design Thinking to Improve Student Innovation.” *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Chesapeake, VA.

APPENDIX A. INSTRUMENT FORMS

The complete test instrument from the first round of testing:

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.
(After 3 minutes, your answers will be saved and you will be automatically sent to the next problem)

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.
(After 3 minutes, your answers will be saved and you will be moved to the next problem)

In the next 3 minutes, write as many SOLUTIONS as you can for the following problem. Garbage can liners often slip down inside of the cans when they are full of garbage.

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

Bike seats are often exposed to the weather and become wet or absorb water, which causes discomfort to the rider.

Solutions:

1) **A plastic cover with elastic around the edge (like a hairnet) that protects the seat from becoming wet.**
2) **A wide fender that folds down to protect the rider from water that splashes from the tire while riding. While not riding, the fender folds up and shields/cover the seat from the weather.**
3) **The seat has ridges that channel water away from the rider and off the surface of the seats.**
4) **Small, removable seat that the rider can take with them while not riding the bike.**

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

People don't like to sit on public toilets. How do we make them more sanitary?
Solutions:

1) **A toilet that automatically sprays disinfectant after every flush.**
2) **Removable toilet seats with a seat washer in the bathroom.**
3) **Seats with multi-layered tissue, one layer is removed after each use.**
4) **Toilet with no seat and people hold on to handrails and squat down.**

For the idea that you picked as best in the previous question, write a pitch to convince people of the benefits of the solution (limited to 700 characters; approx. 9 sentences).

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.

In the next 3 minutes, write as many SOLUTIONS as you can for the following problem:

A local supermarket has to discount their leftover baked goods after they are a day old.

In the next 3 minutes, write as many SOLUTIONS as you can for the following problem:

People often cut across the lawn in places around campus which leaves ugly dead patches in the grass.

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

When people mow their lawns, the grass clippings take up a lot of space after they are done.

Solutions:

1) **A lawnmower that burns the grass clippings after they are cut.**
2) **A grass that doesn't grow longer than the desired length.**
3) **A service that collects clippings and converts them to fertilizer.**
4) **A tank people put in their yard that chemically breaks down the grass into fertilizer.**

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

In classrooms, people often need power in places where there aren't power outlets.
Solutions:

1) **Retractable extension cords built in to the walls.**
2) **Wireless power supply.**
3) **Outlets on a track system so that you can move the outlets around the room.**
4) **Have more charging stations so people don't need outlets.**

For the idea that you picked as best in the previous question, write a pitch to convince people of the benefits of the solution.

The complete test instrument from the second round of testing (with pictures resized to fit the page):

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.
(After 3 minutes, your answers will be saved and you will be automatically sent to the next problem)

In the next 3 minutes, write as many opportunities for innovation (issues that can or should be addressed - NOT solutions) as you can from the image above.
(After 3 minutes, your answers will be saved and you will be moved to the next problem)
In the next 3 minutes, write as many SOLUTIONS as you can for the following problem:

Headphone wires get tangled in people's pockets.

Choose your favorite idea from the previous question and write a pitch about it. Include the following:

- **Describe the problem you are trying to solve.**
- **Describe how your solution fixes the problem.**
- **Be convincing. Persuade people that your solution is a good one.**

(limited to 700 characters; approx. 9 sentences).

In the next 3 minutes, write as many SOLUTIONS as you can for the following problem. Garbage can liners often slip down inside of the cans when they are full of garbage.

Choose your favorite idea from the previous question and write a pitch about it. Include the following:

- **Describe the problem you are trying to solve.**
- **Describe how your solution fixes the problem.**
- **Be convincing. Persuade people that your solution is a good one.**

(limited to 700 characters; approx. 9 sentences).

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

Bike seats are often exposed to the weather and become wet or absorb water, which causes discomfort to the rider.

Solutions:

5) **A plastic cover with elastic around the edge (like a hairnet) that protects the seat from becoming wet.**
6) **A wide fender that folds down to protect the rider from water that splashes from the tire while riding. While not riding, the fender folds up and shields/cover the seat from the weather.**
7) **The seat has ridges that channel water away from the rider and off the surface of the seats.**
8) **Small, removable seat that the rider can take with them while not riding the bike.**

Rank the solutions to the given problem statement - based on the definition of innovation: original and useful ideas that can be implemented successfully. (1=best, 4=worst)

Problem:

People don't like to sit on public toilets. How do we make them more sanitary?

Solutions:

5) **A toilet that automatically sprays disinfectant after every flush.**
6) **Removable toilet seats with a seat washer in the bathroom.**
7) **Seats with multi-layered tissue, one layer is removed after each use.**
8) **Toilet with no seat and people hold on to handrails and squat down.**