122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Making Value for Society
Paper ID #11853

Using an Instrument Blueprint to Support the Rigorous Development of New Surveys and Assessments in Engineering Education

Ms. Jessica Menold Menold, Pennsylvania State University, University Park

Jessica Menold is a second year graduate student interested in entrepreneurship, the design process, and innovativeness of engineering graduates and professionals. She is currently working as a student mentor in the Lion Launch Pad program, where she works to support student entrepreneurs. Jessica is currently conducting her graduate research with Dr. Kathryn Jablokow on a project devoted to the development of a psychometric instrument that will measure the skills, behaviors, and traits of an innovative engineer. Her hope is that this awareness of individual innovativeness levels will enhance engineering professionals and student's innovative skillsets. Jessica is also interested in studying and teaching design thinking methods to students, and is currently working to spread design thinking through mini-workshops across Penn State.

Dr. Kathryn W. Jablokow, Pennsylvania State University, Great Valley

Dr. Kathryn Jablokow is an Associate Professor of Mechanical Engineering and Engineering Design at Penn State University. A graduate of Ohio State University (Ph.D., Electrical Engineering), Dr. Jablokow's teaching and research interests include problem solving, invention, and creativity in science and engineering, as well as robotics and computational dynamics. In addition to her membership in ASEE, she is a Senior Member of IEEE and a Fellow of ASME. Dr. Jablokow is the architect of a unique 4-course module focused on creativity and problem solving leadership and is currently developing a new methodology for cognition-based design. She is one of three instructors for Penn State's Massive Open Online Course (MOOC) on Creativity, Innovation, and Change, and she is the founding director of the Problem Solving Research Group, whose 50+ collaborating members include faculty and students from several universities, as well as industrial representatives, military leaders, and corporate consultants.

Dr. Senay Purzer, Purdue University, West Lafayette

enay Purzer is an Assistant Professor in the School of Engineering Education. She is the recipient of a 2012 NSF CAREER award, which examines how engineering students approach innovation. She serves on the editorial boards of Science Education and the Journal of Pre-College Engineering Education (JPEER). She received a B.S.E with distinction in Engineering in 2009 and a B.S. degree in Physics Education in 1999. Her M.A. and Ph.D. degrees are in Science Education from Arizona State University earned in 2002 and 2008, respectively.

Dr. Daniel Michael Ferguson, Purdue University, West Lafayette

Daniel M. Ferguson is the recipient of four NSF awards for research in engineering education and a research associate at Purdue University. Prior to coming to Purdue he was Assistant Professor of Entrepreneurship at Ohio Northern University. Before assuming that position he was Associate Director of the Inter-professional Studies Program and Senior Lecturer at Illinois Institute of Technology and involved in research in service learning, assessment processes and interventions aimed at improving learning objective attainment. Prior to his University assignments he was the Founder and CEO of The EDI Group, Ltd. and The EDI Group Canada, Ltd, independent professional services companies specializing in B2B electronic commerce and electronic data interchange. The EDI Group companies conducted syndicated market research, offered educational seminars and conferences and published The Journal of Electronic Commerce. He was also a Vice President at the First National Bank of Chicago, where he founded and managed the bank's market leading professional Cash Management Consulting Group, initiated the bank's non credit service product management organization and profit center profitability programs and was instrumental in the breakthrough EDI/EFT payment system implemented by General Motors. Dr. Ferguson is a graduate of Notre Dame, Stanford and Purdue Universities and a member of Tau Beta Pi.

Dr. Matthew W. Ohland, Purdue University

Matthew W. Ohland is Professor of Engineering Education at Purdue University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by over $14.5 million from the National Science Foundation and the Sloan Foundation and his team received Best Paper awards from the Journal of Engineering Education in 2008 and 2011 and from the IEEE Transactions on Education in 2011. Dr. Ohland is Chair of the IEEE Curriculum and Pedagogy Committee and an ABET Program Evaluator for ASEE. He was the 2002–2006 President of Tau Beta Pi and is a Fellow of the ASEE and IEEE.

Using an Instrument Blueprint to Support the Rigorous Development of New Surveys and Assessments in Engineering Education

Abstract:

Many sound methods exist for creating the items or questions that make up educational surveys and assessments. These methods include the use of content experts, reviews of existing instruments, and lists of behaviors and descriptors commonly associated with the construct(s) we wish to assess. Unfortunately, however, item creation sometimes becomes overly dependent upon a researcher's personal attitudes about the construct(s) being tested, or on "borrowing" items from other instruments that may or may not be sound measures of the construct(s) of interest. These risks are particularly likely for new researchers in engineering education, who may have little experience with best practices in social science research.

One way to support best practices in the development of new surveys and assessments is to use an *instrument blueprint* to guide the creation of items, as well as the collection of validity evidence. This paper outlines a process for instrument blueprint creation and content validation to help support best practices in educational assessment. Based on Messick's unified theory of validity[1], the instrument blueprint includes a process for item construction that incorporates multiple resources, including: (1) the views of content experts; (2) research from the relevant domain of interest; (3) reviews of existing instruments; and (4) the expertise of the research team.

This paper uses the development of a new instrument to measure engineering innovativeness as an illustrative example of the blueprinting process. Our new instrument will assess 20 characteristics of innovative engineers as identified by in-depth studies of expert engineering innovators in previous research[2]. This work highlights the positive impact of using a systematic process for item construction to transform current methods of assessment in engineering education.

1. Introduction

Within engineering education assessment, instruments are used for a wide range of activities, from the general evaluation of programs and classes[3, 4] to more specific studies regarding student perspectives and beliefs[5]. Shartrand et al.[4] point to a lack of valid and reliable assessment tools as a weakness in engineering education, and although they were working specifically in the context of engineering entrepreneurship education, we feel this statement can be generalized to engineering education on the whole. According to Downing and Haladyna[6], validity is the most important consideration in test evaluation and refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Haynes et al.[7] go on to warn us that data from an invalid instrument can "over-represent, omit, or under-represent some facets of the construct and reflect variables outside the construct domain".

The use of unreliable and/or invalid instruments in engineering education could lead to the inaccurate measurement of student outcomes and perceptions, incorrect program and class assessments, as well as a general misrepresentation of the current state of engineering education. In this paper, we propose a structured methodology for the initial steps in instrument development, with special attention to item creation. We believe the adoption of this methodology could lead to more rigorous and robust instruments for engineering education assessment. We begin with a standard process model based on the work of Messick[1], Downing and Haladyna[6], and Haynes et al.[7], which we illustrate using a construct from our own work. We begin by reviewing the background research that lead to our current work; we then define twenty characteristics of an engineering innovator and select one to use as an illustrative example of the blueprint process. In the context of this paper, we define an "instrument blueprint" to be the path or process plan that guides the content definition and item creation of an instrument. We present the blueprint to create an item set to be used in initial pilot testing; this process begins with identification and refinement of the construct, moves to the creation and refinement of behavior matrices, and ends with expert review and additional refinement of items.

2. Theoretical Framework

In highlighting the importance of an instrument's validity in the context of score interpretation, Messick notes that the construct validity of score interpretation undergirds *all* score-based inferences.[8] In short, score interpretation is dependent upon the validity evidence collected for the instrument itself, making the rigor of the development process for instruments of critical importance. In this section, we present the theoretical framework of the proposed blueprint; we feel the blueprint presented in this work will aid in the development process for instruments by removing the guesswork in the initial phases of instrument creation, as well as bolstering the validity evidence of instruments.

Content relevance and representativeness are the first steps towards developing a sound instrument and are the focus of the instrument blueprint. Content relevance and representativeness refer to the range and limits of content coverage – i.e., the boundaries of the construct domain to be assessed. Test items are the building blocks of any assessment instrument, and by nature, they specify the content domain of the instrument. In other words, sound instruments are composed of sound items that generate support for the instrument in the collected body of validity evidence. Sound items are grounded in a theoretical framework and are representative of and relevant to the content domain of interest. In the following sections, we review Messick's unified theory of validity[1], which provides a more general overview of the entirety of the instrument development process, as well as the works of Downing and Haladyna[6], and Haynes et al.[7], which provide more specific insights into construct development, test content, and item creation.

2.1 Messick's Unified Theory of Validity

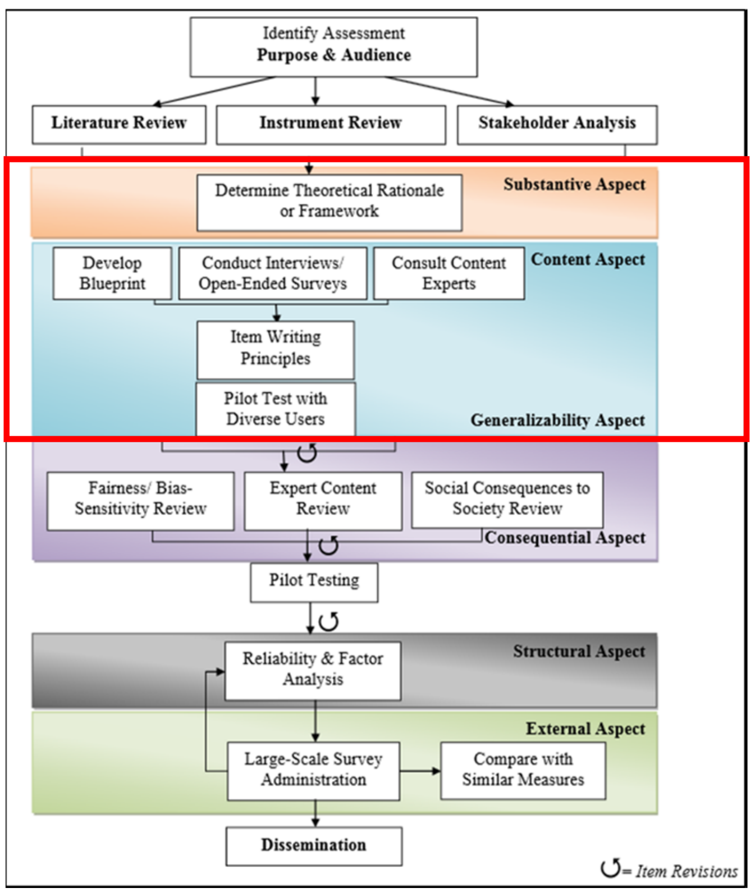
Messick[1] defined validity as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment." According to Messick's theory, validity can be differentiated into six aspects – i.e., content, substantive, structural, generalizability, external, and consequential – with definitions as follows:

- *Content Aspect*: includes evidence of content relevance, representativeness, and technical quality [1, 9]
- *Substantive Aspect*: refers to theoretical rationales for the observed consistencies in test responses[1, 8]
- *Structural Aspect*: appraises the fidelity of the scoring structure to the structure of the construct domain at issue[1, 10]

- *Generalizability Aspect*: examines the extent to which score properties and interpretations generalize to and across population groups, settings, and tasks[1, 11, 12]
- *External Aspect*: includes convergent and discriminant evidence from multitrait-multimethod comparisons, as well as evidence of criterion relevance and applied utility[1, 19, 20]
- *Consequential Aspect*: appraises the value implication of score interpretation as a basis for action, as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice[1].

To illustrate the structure of Messick's theoretical model, Purzer and Cardella[15] transformed Messick's unified theory into a process diagram for instrument creation, as shown in Figure 1. The diagram in Figure 1 outlines a path that instrument developers should follow as they collect validity evidence while creating an instrument. It begins with an identification of purpose of the instrument, as well as the intended audience. Purzer and Cardella then illustrate a multi-faceted review of the state of the field, including a literature review on relevant theories, an instrument review on relevant instruments and scales, and a stakeholder analysis, which could include input from experts in the field of focus. The diagram then moves through the six aspects of Messick's theory of validity and highlights activities and tasks at each step. In this work, we pay specific attention to the steps highlighted in the content and substantive aspects of validity in Figure 1.

Figure 1. Purzer and Cardella's diagram for development of valid measurement instruments[15]



The *content* and *substantive aspects* focus on content relevance and representativeness, or the characteristics of the content domain that is being assessed. The *content aspect* of construct validity serves to specify the boundaries of the construct domain, or the determination of the skills, traits, knowledge, and attitudes that are related to the relevant construct. The content aspect requires that the tasks or behaviors to be assessed are both relevant to and representative of the construct domain. Typically, content relevance and representativeness are assessed by expert professional judgment[1, 6, 7, 9].

With regard to the *substantive aspect* of construct validity, Messick notes: "The substantive aspect refers to theoretical rationales for the observed consistencies in test responses, including process models of task performance, along with empirical evidence that the theoretical processes are actually engaged by respondents in the assessment tasks."[1] In other words, the substantive aspect of content validity adds empirical evidence to the content aspect, in that it supports with test data the theoretical foundations defined in the content aspect of construct validity. This aspect is primarily concerned with the processes representative of a construct and can be evaluated through a variety of exercises, such as "think-aloud" protocols. The point of such exercises is to validate that the tasks or items in the instrument evoke a process/response from the respondents that is consistent with the construct.

Development of an instrument blueprint refers to the identification of behaviors, tasks, skills, and knowledge related to the construct to be tested; the open-ended surveys and consultation with content experts both serve to confirm the blueprint. Once the behaviors, tasks, skills, and knowledge of the related construct are identified, item writing principles[1, 6, 7, 15, 16] are used to guide the formatting and content of the items. This is followed in the process by a pilot test with a sample representative of the final population(s) for whom the instrument is intended. Content creation during the blueprint development stage can be a difficult task, and boiling that content down into a list of cohesive items can be challenging. Downing and Haladyna[6] proposed an ideal process for test item development, as discussed in the following section.

2.2 Item Development

As Downing and Haladyna[6] note: "The gathering of supporting evidence for validating a specific test use or interpretation must begin with a careful and systematic approach to the task of creating the test items." This statement highlights the importance of using a standard method for content creation, as well as item development. Haynes et al.[7] suggest that content creation for a specific instrument begin with content definition, or, in other words, with defining the scope of the construct domain.

Carefully defining and differentiating the construct to be evaluated is an important first step in instrument development that should not be overlooked. Content definition refers to the selection of the survey/instrument domain and associates the construct that is to be measured with the test specifications and items. This first form of validity evidence clearly defines the boundaries of the assessment. Job-task analysis is often suggested as key in defining the content of a survey or instrument and simply involves evaluating the construct in its "natural environment". Content may also be defined from extant literature or from existing theories that have been accepted by some panel of experts in the field[6, 7].

The next step in item development suggested by the literature[1, 6, 7, 8, 9] is the construction of a detailed roadmap of the instrument – i.e., the test specifications. The test specifications lay out a guide to ensure that items are both relevant and representative of the content domain being assessed – i.e., that the items are related to the knowledge, skills, attitudes, and behaviors to be assessed[1]. The test specifications define the type of content within the assessment, as well as specifying the size of each content category to be present in the survey instrument[6]. In future sections, we present an example of test specifications using our own research in engineering innovativeness.

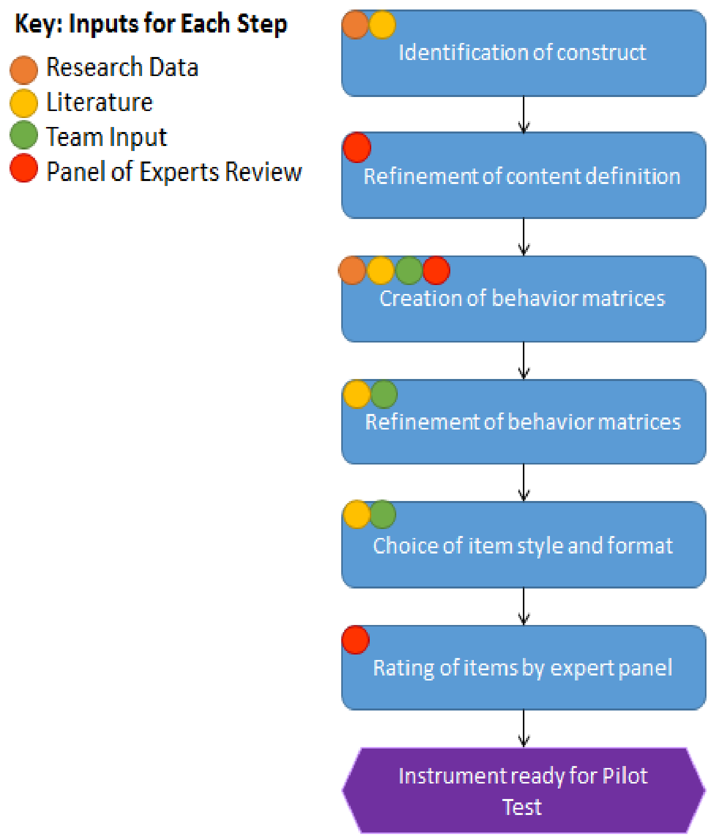
When creating items for an instrument, the content is typically based on expert experience, textbooks, or a thorough literature review. However, in order to verify that the content reflected in the items is representative and relevant to the domain of assessment, it is commonplace to organize a panel of experts to evaluate the item set. Typically, this panel of experts is composed of professionals with significant knowledge and/or experience in the domain being assessed. The panel is also briefed on what is to be assessed – i.e., the panel is explicitly told the definition(s) of the construct(s) being assessed and is instructed to evaluate the items with reference to this definition(s). This helps to avoid dissention amongst the experts on such things as content definition or theoretical framework, and helps maintains focus on the relevance and representativeness of the items[1, 6].

2.3 Proposed Blueprint for Item Creation

Taking into account the theoretical frameworks described above, and paying specific attention to item development, we present an addendum to Purzer and Cardella's instrument development diagram in Figure 2. This new diagram is a detailed representation of the tasks and steps required in the substantive and content aspects highlighted in Figure 1 (in the red box), and is the basis for the proposed blueprint process for content definition and item creation. The generalizability aspect, although included in the highlighted box of Figure 1, begins once pilot testing is complete and thus is not expanded upon in Figure 2.

The purpose of this model is to provide clarity on the specific tasks and actions needed to develop test items with high content validity. The steps presented in Figure 2 represent the tacit knowledge pertaining to item development gained by this research team in the creation of a new instrument to measure engineering innovativeness. The inputs for each step are indicated by the colored circles in the top left of each step box, which may take the form of research data, extant literature, team input, and expert reviews. Inputs for our own work included previous research data (i.e., interviews with engineering innovators); literature related to traits, behaviors, knowledge, and skills related to innovativeness; input from the research team; and input from a panel of experts within engineering innovativeness.

Figure 2:Blueprint for Content Definition and Item Creation



3.0 Illustrative Example

3.1 Project Context

The work described here is part of a long-term research agenda aimed at identifying and assessing the key factors of engineering innovativeness[19]. Specifically, our work involves the development of a socially constructed definition of engineering innovativeness, the development and validation of a new instrument to measure engineering innovativeness, and the creation of a benchmark database of engineering innovativeness among engineering students and practitioners.

To date, we have completed a series of studies to develop and confirm a socially constructed set of key engineering innovativeness factors. After interviewing 53 engineering innovators about their experiences and qualitatively analyzing the interview data, we identified twenty unique characteristics of engineers who had demonstrated extraordinary innovative behavior[17]. This finding was corroborated by a separate focus group study[17] and a modified Delphi study with 150 engineering innovators drawn from academic, corporate, and entrepreneurial organizations[2]. An in-depth literature review was carried out to identify how constructs related to the 20 characteristics are currently assessed (when such an assessment exists)[18], leading to our current work in developing and validating a new instrument.

For the purposes of this paper, we will select *one* of the 20 key characteristics of an engineering innovator and illustrate the proposed process model for creating and validating items for a new assessment instrument. Specifically, we will walk through the steps used to derive items for the "curiosity" construct, beginning with domain definition[19].

3.2 Domain Definition and Construct Identification

The first step in identifying the content domain and construct definition of "curious" involved the input of engineering innovators as content experts. Through interview analysis, an initial definition of the construct was derived; this definition was then subjected to further validation through a Delphi Study[20]. The Delphi study was conducted by recruiting experts in the field of engineering innovativeness. Experts were asked to rate their agreement with the current definition, suggest possible changes to the definition content, as well as rank each construct's relative importance in a three-phase model of innovation (see Figure 3). The results of the last ranking helped guide the behavior matrix creation, which will be discussed in section 3.3. Final definitions for each of the twenty characteristics are presented in Appendix A.

The final definition for "curious" that resulted from this study was "actively challenges themselves to learn or know more about something". The steps highlighted above map to the first two blocks in Figure 2. The use of experts in the derivation of constructs, as well as the refinement of the content domain, was a key aspect of this research. These steps helped ensure that our constructs were differentiable and unique from past work, as well as to ensure the relevance and representativeness of the current state-of-the-art in engineering education, with specific regard to engineering innovativeness work.

3.3 Test Specifications and Behavior Matrices

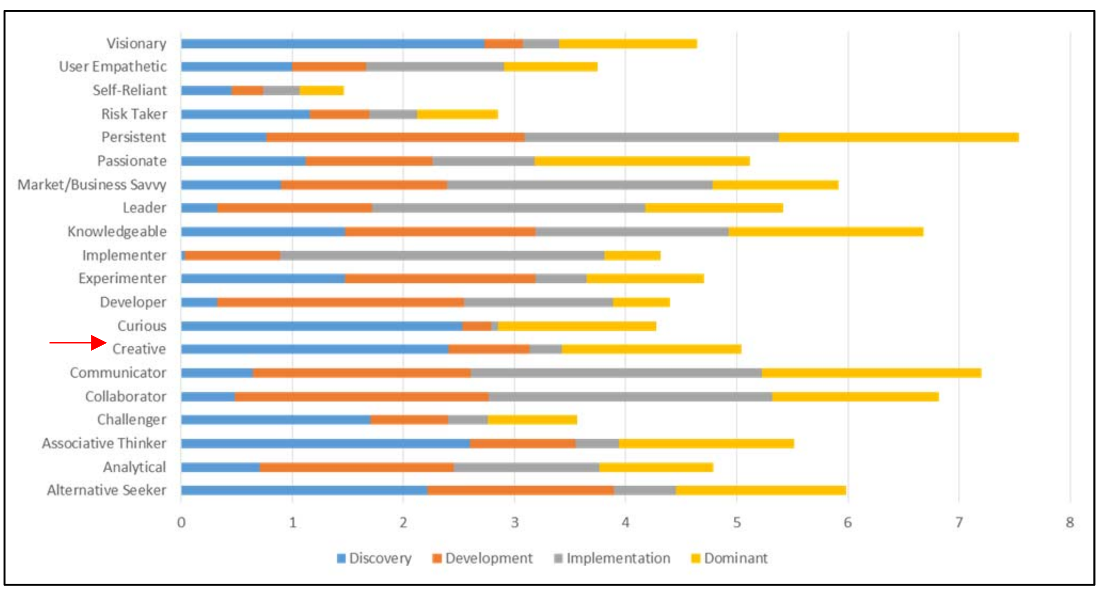
After the definition of the construct domain, test specifications for the instrument were designed. Typically, test specifications highlight the content representation within the instrument. For example, a chemistry test may focus 50% of the items on content related to organic chemistry and 50% of the items related to inorganic chemistry. The purpose of the test specifications is to provide item writers with a guide for item construction and is used to dictate how many items are "dedicated" to each piece of the content domain of interest. For our purposes, the test specifications are an even breakdown across all twenty characteristics. In other words, the items of the final instrument should evenly represent each of the twenty characteristics, so if our instrument has 100 items, each characteristic will have five items related to its content domain. This breakdown was derived based on the initial interview analysis, as well as the results of the Delphi study.

As mentioned previously, an important aspect of the Delphi study was ranking each characteristic's relative importance with respect to the three-phase model of the innovation process. This three-phase model is presented in Figure 3. Although the process is presented linearly, we understand the process to be iterative in nature; this innovation model was also derived from data collected during initial interviews with engineering innovators[17].

Figure 3: Three-Phase Innovation Process Model

Experts were asked to rank the level of importance of each of the 20 characteristics in each of the phases of the innovation process or as dominant throughout the entire process. Participants in the Delphi study were asked to select up to 7 characteristics they felt were most important in each of the three phases, as well as 7 characteristics they felt were necessary or dominant throughout the entire innovation cycle. Results of these rankings are shown in Figure 4. Characteristics more important in the discovery phase (blue bar) were more frequently chosen by Delphi participants as among the top 7 characteristics and will have a larger blue bar in the figure below. This is true for each phase, with blue representing the discovery phase, orange representing the development phase, grey representing the implementation phase, and yellow representing dominant throughout the entire process. For example, Curious was frequently rated as necessary in the discovery phase of the innovation process, and as a result, has a larger blue bar than other characteristics in this phase. These lines represent the frequency with which participants selected characteristics as necessary or important; as a result, the overall length of each bar is an indication of the frequency with which a characteristic was selected at all.

Figure 4: Rankings of Characteristics with Respect to Innovation Process



After reviewing and editing the constructs based on the data from the Delphi study, our team began to review past research data, relevant literature, and related instruments for each of the 20 constructs. Although general item writing principles[1, 16] exist, they do not provide a structure or process plan for item creation. As such, we constructed behavior matrices for each construct based on previous research, interview data, and the results from the Delphi study. Behavior matrices refer to a table or matrix composed of descriptive texts, derived from interview and Delphi data, literature, and existing instruments, that refer to and define the content domain for the construct of interest. These matrices form the basis of the initial item pool for each characteristic. The team derived approximately 15-20 behaviors per characteristic and sorted them based on their theoretical underpinnings. The team collected evidence from over 27 existing innovativeness instruments, and the items (as well as underlying theories) were the basis of some of the behavior creation[20].

In order to generate a pool of behaviors that accurately represented each of the 20 characteristics, the behavior matrix was first broken down by phase. Then, relying on descriptive data from the interviews and Delphi results, behaviors were derived and sorted into each of the phases, per characteristic. For example, as previously stated, Curious was rated as being more important in the discovery phase; as a result, more emphasis was placed on creating descriptive behaviors of a curious individual that related to the discovery phase. This framework was used as a technique for item creation, not to ever be used in the final instrument. As an example of this type of behavioral description for Curious, we find the following displayed in Table 1: "Asks questions to get to the root of the problem." This describes a behavior related to problem definition, a key task early on in the innovation process, specific to the discovery phase. The purpose of these matrices was to provide the research team with a more structured approach for item creation; an example of the full behavior matrix for Curious is shown in Appendix B. After compiling matrices for each of the 20 characteristics, the behaviors were ranked by each research team member, and the top 7-10 behaviors from each characteristic were selected for review by a content validity panel. Behaviors are used here as the basis for items for the pilot study; from this point onward, these behaviors will be referred to as items of the initial instrument.

3.4 Content Validity Panel

Prior to sending selected items to the content validity panel, it was necessary to select an appropriate item format and type. We consulted expert psychometricians, as well as existing literature, to guide the selection of item format and style[16, 21]. As discussed earlier, and as shown in Figure 2 as *rating of items by expert panel*, the purpose of the content validity panel is to review the initial items derived from the behavior matrices. Panel members were recruited based on their content knowledge of engineering innovativeness. The content validity panel was composed of experts from both industry and academia, because the proposed final use of our instrument will impact both academic and corporate populations (students and professionals). Panel members were asked to rate their agreement with each of the behaviors on a five-point Likert-type scale, with choices ranging from strongly disagree to strongly agree. Panel members were provided with the characteristic definition of each of the constructs (derived from interview and Delphi results) and prompted as follows: "Given the definition of [characteristic] above, please rate your agreement with each of the behavioral descriptions listed below." The panel was composed of 23 members, and after two weeks, the survey was closed and results were analyzed.

For each item, the average rank and content validity ratio (CVR) were calculated. Rank revealed the strength of each item, while CVR revealed the degree of agreement per item. The exact method to calculate the Content Validity Ratio actually starts with judges evaluating the item as being essential, useful, or not necessary. For the purposes of our work, a rating of 4 or 5 on the

Likert-type scale was considered essential, 3 was considered useful, and 2 or 1 was considered not necessary. The CVR is then computed as follows: (n - (N/2)) / (N/2), where n is the number of judges rating the item as essential, and N is the total number of judges.

The resulting CVR value is evaluated against pre-established critical values.[22] Items with a CVR value below the critical value would be eliminated. In our case, items that did not meet the minimum CVR value for 20 raters, which is .42[22], were eliminated. If there was a fifth/sixth place tie, the CVR was used to judge which item had more agreement. The top five items from each characteristic were chosen for the next round of pilot testing. An example of ratings and CVR scores for Curious, as well as rankings of the items, is shown in Table 1 below.

Table 1: Average Ranked and CVR Score from Content Validity Panel

Curious			
Items	***Avg Score***	***CVR Score***	***Rank***
Has an inquiring mind	4.7727273	0.9090909	**1**
Asks questions in order to get at the root of the problem	4.7272727	1	**2**
Actively challenges themselves to learn or know more about something	4.6363636	0.9090909	**3.5**
Wonders why things are done one way and not another	4.6363636	0.9090909	**3.5**
Asks good questions and listens to others	4.6363636	0.8181818	**5**
Pursues unexplored solution spaces	4.5909091	1	**6**
Is not afraid to ask how something works	4.5	0.9090909	**7**
Looks for new discoveries	4.4545455	0.7272727	**8**
Constantly trying things out to see what might happen	4.3181818	0.7272727	**9**
Enjoys taking things apart and learning how and why they work	4.3181818	0.6363636	**10**

4.0 Conclusions, Limitations, and Future Work

The purpose of this paper was to highlight an efficient method for item construction, as well as to suggest a more standardized method of instrument construction. We feel that following the outlined method above could lead to more rigorous and robust research in the field of engineering education. The creation of our own instrument certainly benefited from using the above process; if we had used less rigorous methods, we feel two opportunities would have been missed. First, by incorporating the input from experts at such regular intervals, we not only bolstered the validity of our instrument, but we were able to develop relationships with potential collaborators in future research work. These relationships will be key later in the instrument development process, when gathering large samples from a variety of sources for the pilot study is necessary to evaluate the reliability and validity of our instrument. The second opportunity that may have been missed if another method had been used was the failure to gather validity evidence early in the instrument development process. Gathering validity evidence is a time consuming endeavor that occurs over the lifetime of the instrument.[23] As such, it is necessary to collect validity evidence whenever possible, especially in the early years of an instrument. If we had not used the outlined process, the collection of validity evidence in these initial phases of development would have been haphazardly done or the opportunity to collect evidence might have been missed entirely.

The blueprinting process for item creation creates a path and sets guidelines for instrument creation. By utilizing a set or standard method, the need to re-do or re-create items may be eliminated, and the initial item development of an instrument could become a streamlined process. We feel the process outlined above will help support educational assessment through the use of standardized best methods for blueprint creation and content validation.

References

1. Messick, Samuel. "Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning." American Psychologist 50.9 (1995): 741.
2. Purzer, Senay, and Kathryn Jablokow. "Collaborative Research: Identifying and Assessing Key Factors of Engineering Innovativeness."
3. Budny, Dan, William LeBold, and Goranka Bjedov. "Assessment of the Impact of Freshman Engineering Courses*." *Journal of Engineering Education* 87.4 (1998): 405-411.
4. Shartrand, Angela, et al. "Assessing student learning in technology entrepreneurship." *Frontiers in Education Conference*, 2008. FIE 2008. 38th Annual. IEEE, 2008.
5. Li, Qing, et al. "Development of an instrument to measure perspectives of engineering education among college students." *Journal of Engineering Education* 97.1 (2008): 47-56.
6. Downing, Steven M., and Thomas M. Haladyna. "Test item development: Validity evidence from quality assurance procedures." *Applied Measurement in Education* 10.1 (1997): 61-82.
7. Haynes, Stephen N., David Richard, and Edward S. Kubany. "Content validity in psychological assessment: A functional approach to concepts and methods." *Psychological assessment* 7.3 (1995): 238.
8. Embretson, Susan, and Joanna Gorin. "Improving construct validity with cognitive psychology principles." *Journal of Educational Measurement* 38.4 (2001): 343-368.
9. Lennon, Roger T. "Assumptions underlying the use of content validity." *Educational and Psychological Measurement* (1956).
10. Loevinger, Jane. "Objective tests as instruments of psychological theory: Monograph Supplement ." *Psychological reports* 3.3 (1957): 635-694.
11. Cook, Thomas D., Donald Thomas Campbell, and Arles Day. *Quasi-experimentation: Design & analysis issues for field settings.* Vol. 351. Boston: Houghton Mifflin, 1979.
12. Shulman, Lee S. "Reconstruction of educational research." *Review of Educational Research* (1970): 371-396.
13. Campbell, Donald T., and Donald W. Fiske. "Convergent and discriminant validation by the multitrait-multimethod matrix." *Psychological bulletin* 56.2 (1959): 81.
14. Cronbach, Lee J., and Goldine C. Gleser. "Psychological tests and personnel decisions." (1965).
15. Purzer, Senay and Cardella, Monica. Instrument Development Model: A Map based on Messick's Unified Theory of Validity, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
16. Kline, Paul. *A handbook of test construction: Introduction to psychometric design*. Methuen, 1986.
17. Ferguson, D., Purzer, S., Ohland, M., Jablokow, K., and Menold, J. 2014. “Characterizing engineering innovativeness through a modified Delphi study,” Proc. of the 2014 Frontiers in Education Conference, Spain.
18. Menold, Jessica, Jablokow, Kathryn, Ferguson, Daniel, Purzer, Senay, and Ohland, Matthew. "A Critical Review of Measures of Innovativeness."

19. Purzer, Senay, and Kathryn Jablokow. "Collaborative Research: Identifying and Assessing Key Factors of Engineering Innovativeness."
20. Ferguson, Daniel M. "How engineering innovators characterize engineering innovativeness: A qualitative study." (2013).
21. Suen, Hoi K. *Principles of test theories*. Routledge, 1990.
22. Lawshe, Charles H. "A quantitative approach to content validity1." *Personnel psychology* 28.4 (1975): 563-575.
23. Joint Committee on Test Standards of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*.

Appendix A: Definitions of Twenty Constructs

Characteristic Name	New Characteristic Definitions based on Round 1 Participant Comments
Alternatives Seeker	**Actively searches for multiple choices or solutions or new non-obvious options to make something better.** Looks beyond what they or others know. Looks outside of their own area of specialty to find solutions.
Analytical	**Separates something into component parts or constituent elements.** Uses mathematical or other modeling techniques to synthesize problems. Simplifies complex systems into sub-systems or elements. Attentive to accuracy and detail. Able to think critically.
Associative Thinker	**Joins or connects together ideas or facts from different domains or experiences**. Sees connections that help them to explore the solution space. Does systems thinking. Able to transpose experiences and observations across seemingly unrelated domains.
Challenger	**Questions the current state of things.** Driven to find a better way. Challenges the status quo and thinks of ways to improve and make better the current process or product. Is skeptical and reluctant to accept conventional thinking or what may appear obvious.
Collaborator	**Actively networks with people in or supporting the project.** Integrates the knowledge of others into a better solution than from any one individual. Incorporates ideas and strategies of others that may differ from their own. Brings together people with a diversity of knowledge and utilizes the collective knowledge to solve a complex problem.
Communicator	**Explains the idea, the concept, and the opportunity by speaking, writing, gestures or use of pictures or diagrams**. Tells a story to advocate for the idea. Makes the idea easy to understand to sell the idea.
Creative	**Invents a new product, process or concept that has value to a community. Applies existing technology in a new way to solve a problem.** Takes a unique approach to solving a problem, sharing information, connecting ideas, or exploring options.
Curious	**Actively challenges themselves to learn or know more about something.** Seeks or pursues unexplored paths. Is eager to learn and experience something. Asks good questions and listens to others. Is inquisitive and purpose driven.
Developer	**Enables self and others by breaking down barriers and obtaining sufficient resources to move something ahead.** Brings ideas to life and demonstrates them. Turns ideas into real world solutions. Focused on making "something" better in some way.
Experimenter	**Performs a series of actions and carefully observes their effects in order to learn about something.** Tries an idea out in situations to help develop the idea. Conducts tests, fails, learns, and retests. Fails forward fast.
Implementer	**Takes an idea from development into an end product.** Turns ideas into workable systems that last a full life cycle in implementation. Takes the idea to a usable conclusion. Completes something despite obstacles or barriers. Accomplishes something tangible.
Knowledgeable	**Possesses information, understanding, or skill that spans a significant number of different subject areas.** Is skilled in independent learning. Possesses knowledge that is both broad and deep. Is technically excellent in their field.

Market/ Business Savvy	**Possesses practical understanding or knowledge of [business or market] and able to use this knowledge to identify unmet needs**. Sees and relates the idea to business value and holds an idea until the moment is right. Focuses on methods of profitably satisfying customer needs.
Passionate	**Expresses strong emotions or beliefs about something.** Is enthusiastic and energetic about the idea. Works on something that inspires them because it aligns with their created and given strengths.
Persistent	**Continues to do something even though it is difficult or other people want you to stop.** Focuses on reasonable solutions to problems in alignment with the value of solving them. Continues beyond the usual or expected effort. Drives a potential innovation through to either implementation or validation of not being feasible/viable.
Risk Taker	**Doesn't think failure is bad, but that failure provides for learning**. Accepts the possibility of being wrong. Acts when a situation has unknowns and uncertainties. Takes calculated risks that balance rewards/risks.
Self-Reliant	**Confident in own abilities and able to do things for yourself.** Seeks other resources to close the gap on what they are not good at. Confident enough learner and practitioner to ask for help where they are weak. Personally motivated to define a problem and pursue a solution.
Leader **[new name]**	**Inspires other individuals and facilitates achieving a key result or a group of aligned results**. Builds an environment to make others effective at working together. Empowers others to deliver against the common goal. Envisions an end goal and influences others to rally together towards the goal.
User Empathetic	**Understands the feelings, thoughts or experiences of another person/group.** Is ethical and humble regarding customer or stakeholder needs. Knows what is important to the stakeholder audience and/or customer. Investigates the full innovation life cycle - not just the inception and initial implementation.
Visionary	**Has ideas about what could/should happen or be done in the future based on an understanding of user needs.** Able to see how a solution to a current problem can be fully implemented in the future. Sees strategic long term value of idea to end user, customer, and organization.

Appendix B: Item Behavior Matrix for Curious

Characteristic	Discovery Behaviors	Development Behaviors	Implementation Behaviors	Dominant Behaviors
Curious: Actively challenges themselves to learn or know more about something. Seeks or pursues unexplored paths. Is eager to learn and experience something. Asks good questions and listens to others. Is inquisitive and purpose driven.	Constantly wondering why things are done one way and not another	Explores features within the solution space	Curious and interested in learning about a variety of markets	Have a love of learning and a deep to desire to continuously learn more
	Eager to learn from others more experienced with the problem	Constantly trying things out to see what might happen.	Curious about various advertising methods and business plans	Inquisitive, have an inquiring mind and are never satisfied with an answer, want to know more
	Aware of the limitations of their own knowledge base			Always asking questions, particularly why questions
	Not afraid to ask how something works			Enjoying reading and staying on top of what's new in their technical field
	Have a desire to make new discoveries			Looks forward to learning challenging concepts
	Have no or very little preconceived notions about the problem			Regardless of the topic they are eager to learn more about it
	Enjoy taking things apart and learning how and why they work			Desire to constantly be learning new things
	Enjoys searching for solutions to complex problems			Motivation towards increasing their knowledge or skillset
				Absorbs facts and details even if they are not immediately useful