Board 36: A Concept Inventory for Functional Reasoning in Engineering Design

Prof. Michael J Scott, University of Illinois, Chicago

Michael J. Scott is Associate Professor of Mechanical Engineering at UIC, Director of Graduate Studies for the Department of Mechanical and Industrial Engineering, and Director of the Interdisciplinary Product Development Program.

A Concept Inventory for Functional Reasoning in Engineering Design

Abstract

A concept inventory is being developed for functional reasoning in engineering design. A Delphi process used a panel of experts to identify and prioritize candidate concepts for inclusion in the concept inventory over several rounds of feedback. Concepts were selected on the basis of their relative difficulty for learners to understand and their perceived importance by the panel. For the top concepts, multiple choice questions are under development with distractor answers that pointed to expected misunderstandings. These questions are to be tested with groups of students from an introductory design course that covers functional reasoning, and the concept inventory will be created from the validated questions.

The development process diverged from a model process used previously to develop a concept inventory for thermodynamics. The peculiarities of this concept inventory development process and possible lessons for such development in general will be discussed.

1 Introduction

This is a report of work in progress on a project to develop a concept inventory for functional reasoning in engineering design by means of a Delphi process. Functional modeling procedures are often taught in undergraduate design classes as a part of reverse engineering or as an early step in creating new designs. However, there is no accepted means of assessing whether students understand the underlying concepts when applying these procedures, and indeed there is no agreed-upon articulation of the concepts themselves. The usual goal of a concept inventory is to help educators to determine if their students understand the concepts of a field, rather than simply being able to implement procedures. In the case of functional reasoning, the concept inventory may help to define its conceptual underpinning. The impact of a concept inventory on education in design could be to initiate changes in the way that function is taught.

2 Background: Delphi method, concept inventories, and functional reasoning

The exposition of this project depends upon a basic understanding of the definitions of functional reasoning in engineering design, concept inventories, and the Delphi method. These three ideas are thus reviewed briefly.

2.1 Functional reasoning in engineering design

The functional reasoning considered here is the description of devices or systems by their functions in the tradition that follows the work of Pahl and Beitz [1], in which a function is defined as "an abstract formulation of (a) task, independent of any particular solution". Functions in engineering design are specific, abstract, and independent of any particular embodiment, even when a functional representation is built around a particular device. The teaching of systematic design methods tends to gloss over the definition of the term *function* and proceed to methods for producing functional representations, for which there is more variety. (Even when textbooks begin with a dictionary definition, such as Dym et al. [2], the main exposition is in terms of methods.) Nagel et al. [3], for instance, identify and compare seven different methods: the Glass Box Method, Function Analysis System Technique or FAST, flow-based Systematic Processes, enumeration, the Zen approach, reverse engineering, and function-means trees. After reviewing these methods and the representations they produce, they comment: "The biggest challenge is getting the idea of function across to our students...The challenge is teaching 'function' as an abstraction, not as a means to modelling functionality."

Researchers have offered different views of what this idea of function is or should be. Erden et al. [4] give an excellent review of 18 different approaches (see, in particular, their Table 1), which they categorize with respect to six different considerations: ontology, semantic definition, formalism, context, decompostion/verification, and implementation. Pahl and Beitz's [1] systematic design, probably the closest thing there is to a canonical reference for undergraduate education, is classified as having a device-centric rather than process-centered ontology, which is the case for 17 of the 18 listed approaches (the notable exception being the Direct FM approach of Umeda and Tomiyama and co-authors). On the other hand, Pahl and Beitz are identified as the only method that covers three types of semantic definition: subjective/purposive, verb/noun linguistic description, and input-output/tranformation of flow. This may come as a surprise to the attentive undergraduate who learned that these three ideas — that functions can describe the purpose of a system, that functions are expressed as verbs with objects, and that functions transform flows of energy, signal,[1] and material — are fundamental to the conception of function. Yet none of the other 17 approaches surveyed covers all three, and several are classified as defining functions as having a direct mapping to components [5, 6], when Pahl and Beitz are not. Only one group [7, 8, 9] is credited with developments in process-oriented ontologies after original work by Forbus [10]. In proposing a framework that distinguished purpose from action functions, and function from behavior, Deng [11] noted that "Function is an overloaded term." It would appear to still be the case that, as Umeda and Tomiyama [12] opined over twenty years ago: "(function) has no clear, uniform, objective, and widely accepted definition."

2.2 Concept Inventories

The use of *concept inventories* [13, 14, 15] to assess students' conceptual understanding is gaining popularity in engineering and related fields. The best-known concept inventory is probably the Force Concept Inventory [16] in physics; see Epstein [17] for a good discussion of the importance

---

[1] Some Delphi panelists objected to the word *signal*; this is discussed below.

of the FCI in awakening educators to the notion that conceptual understanding might not be well-judged by traditional exams. Engineeering design classes often use a project format, so conceptual understanding is not assessed simply because holistic assessment of student project work does not usually include assessment of conceptual understanding of functions and functional reasoning.

Concept inventories are validated tests, usually in a multiple-choice format, that not only check for understanding of a given concept, but also provide "distractors" that can help show if students hold particular misconceptions. This is explained by Pellegrino et al. [18]:

> Concept Inventories provide good examples of the use of conceptual odels of understanding in instructional domains to systematically generate sets of test questions (the cognition and observation vertices of the assessment triangle). In general, Concept Inventories are intended to measure students' deep conceptual understandings within a relatively narrow domain and they share a number of common features. Typically, they consist of multiple choice questions whose distractor options were developed by first asking students open ended questions and analyzing the student responses obtained. The questions are frequently based on science and engineering education research, including research on misconceptions and common student errors. Many inventories have developed explicit links between incorrect multiple-choice distractors and specific misconceptions.

Concept inventories have been developed in thermodynamics [19], statics [20], and materials [21], among other engineering fields, and have been collected in repositories such as ciHUB [14].

2.3 The Delphi Method

This work developing a concept inventory for functional reasoning follows the recommendation of Streveler et al. [22] to begin by convening a *Delphi panel* to generate a list of concepts for inclusion in the inventory. A Delphi panel is a group of experts in a given field who interact anonymously through several rounds of questionnaires. Anonymity is desirable so that panelists discuss the ideas on their merits alone and do not defer to the perceived authority of any other panel members. The Delphi panel here consisted of 14 academics with ranks ranging from Assistant Professor to Professor Emeritus in mechanical engineering or industrial engineering. Eleven of the panel had were employed in and had doctoral degrees from North America, and the other three were from Europe or Australia. All had taught design classes, and one of the panelists had just moved to an academic position after over two decades in industry.

In Round Zero of the Delphi process, panel experts are asked to suggest concepts for possible inclusion in the inventory. It is interesting that the functional reasoning Delphi panel was much more reticent to suggest concepts than the panel that was convened by Streveler et al. [22] to produce a concept inventory for thermodynamics. Subsequent rounds of the Delphi process are meant to allow the panel to converge on a subset of the initial list by having the experts rate the candidate concepts on the basis of their importance and understanding. Experts also provide comments when giving their ratings, and experts see the aggregate results from Round One before giving ratings in Round Two. The idea is to seek convergence to agreement on the ideas over the successive rounds.

3 Delphi method results

The Delphi panel responded to two successive rounds of ranking of concepts. Given the small number of active participants in Round Zero, the panel was augmented to a total of 16 experts, of whom 13 gave complete responses over two rounds of questions and one completed Round One and partially completed Round Two. The collection of responses took several months for each round.

Both Round One and Round Two questions were structured as follows. For each concept under consideration the following selections were offered:

1. Select one of the following:
    (a) This concept should be included in this list: Yes/Maybe
    (b) This concept should NOT be included in this list

   If the negative choice is made, no further questions are asked.

2. Select the option that best describes the level of **understanding** of this concept:
    (a) Almost no students understand it
    (b) About 25% of students understand it
    (c) About half of students understand it
    (d) About 75% of students understand it
    (e) Almost all students understand it

   The level of understanding is a measure of the difficulty of the concept.

3. Select the option that best describes the level of **importance** of this concept:
    (a) Unimportant
    (b) Somewhat important
    (c) Important
    (d) Very important

4. (Round One) Select the appropriate level(s) for this concept:
    (a) Undergraduate
    (b) Graduate

   (Round Two) Indicate if this concept is only appropriate at a graduate level.

Table 1: Paired t-test for comparing the standard deviations of average importance and average understanding between rounds one and two

	Average Importance	Average Understanding
Round One Std. Dev. Mean	0.6600	24.520
Round Two Std. Dev. Mean	0.7086	21.540
Difference	-0.0486	2.980
p-value of t-test of Mean Difference = 0	0.125	0.003

5. After the questions there is an optional Comments field. Experts used this field to suggest new concepts, modifications to concepts, and combinations of existing concepts.

Comments and averaged results from Round One for each concept were provided to the panel together with the questions during Round Two. Note that the question about whether the concept was appropriate at the graduate or undergraduate level was modified from Round One to Round Two; this question turned out to provide little value in either round.

Round One results for item 4, the question of whether the concept was appropriate for graduates or undergraduates, were considered unreliable due to the format of the questionnaire. The intent of the question was to identify concepts which are *only* appropriate at a graduate level, but panelists were given two checkboxes, one for the undergraduate level and one for the graduate level. This led to possible confusion in how panelists interpreted the question. For example, if a concept was necessary at an undergraduate level, it should follow that the concept is still appropriate at a graduate level. What does it mean to check just the undergraduate box rather than both the undergraduate and graduate boxes? In an attempt to rectify this confusion, Round Two's questionnaire was modified to provide only a single optional selection of "graduate level only".

3.1 Convergence of results between rounds

Fifty-six candidate concepts were included in Round One, and six were removed from consideration in Round Two because a majority of respondents declared that the concept should not be included in the list. The remaining fifty were carried over to Round Two.

In a Delphi process, the experts' feedback is expected to converge to a consensus over successive rounds [23]. To assess the convergence between Round One and Round Two, the paired t-test at 95% confidence interval is used to see whether there is a statistically significant difference between standard deviations of average understanding across all questions before and after the second round of Delphi process. The same test is used to compare the difference between standard deviations of average importance for the first and second rounds. Results of these two tests are provided in Table 1. These results are aggregated across all candidate concepts; a line-by-line examination of the standard deviations for each concepts finds no particular pattern.

Table 2: Pearson's *r* of correlations between keep, average importance, and average understanding, round two

	Keep	Avg. Imp.
Avg. Imp.	0.502	
Avg. Und.	0.504	0.528

No significant difference for the standard deviation of the average importance of concepts was found. The *p*-value of the first test implies that experts generally preferred to stick to their first-round opinions about the importance of the concepts; in other words, experts were not inclined to adjust their sense of the importance of the concepts in response to other experts' opinions. In the second test, there is a significant reduction in the standard deviation of the average understanding, which indicates some convergence towards consensus among the experts. This can be interpreted as a willingness by the experts to modify somewhat their views of how difficult the concepts were when offered the experience of other experts. The item-by-item results also show a general lack of convergence. Of 48 concepts carried from Round One to Round Two, only 14 showed convergence on both importance and understanding, 5 showed divergence on both, and the remainder were split.

For neither criterion did it appear that a third round of questions was likely to achieve greater consensus. The ratings of importance did not converge at all from Round One to Round Two, and while the ratings of understanding did show convergence at a level of significance of 0.003, the average standard deviation moved only from 24.5 to 21.5.

3.2 Selection criteria for concepts

According to Streveler et al. [22], the list of concepts for inclusion in the concept inventory should be those that are deemed by the panel to be both important and difficult to understand. These questionnaires included two questions in addition to those covering importance and understanding. First, there was the opening question asking the experts simply to declare whether the concept should be included, which we referred to as the "Keep" question. As mentioned above, the Keep question was used to eliminate six concepts after Round One.

Correlations between the Keep and Importance results, and between the Importance and Understanding results, are shown in Table 2. Since the ratings on the Keep question correlated highly with the panel ratings on the Importance question, the Keep value did not provide additional information useful to select the proper concepts for inclusion in the concept inventory. The results are interesting, however, in that they show that the experts apparently ignore the difficulty of a concept when asked whether a concept should be included in the inventory, and rely on their judgment of its importance.

The Importance and Understanding results are also positively correlated. The Pearson correlation value is 0.528 and the P-value is less than 0.0005, which is statistically significant. This was a surprising finding. A higher value of importance means a concept is a better candidate for inclu-

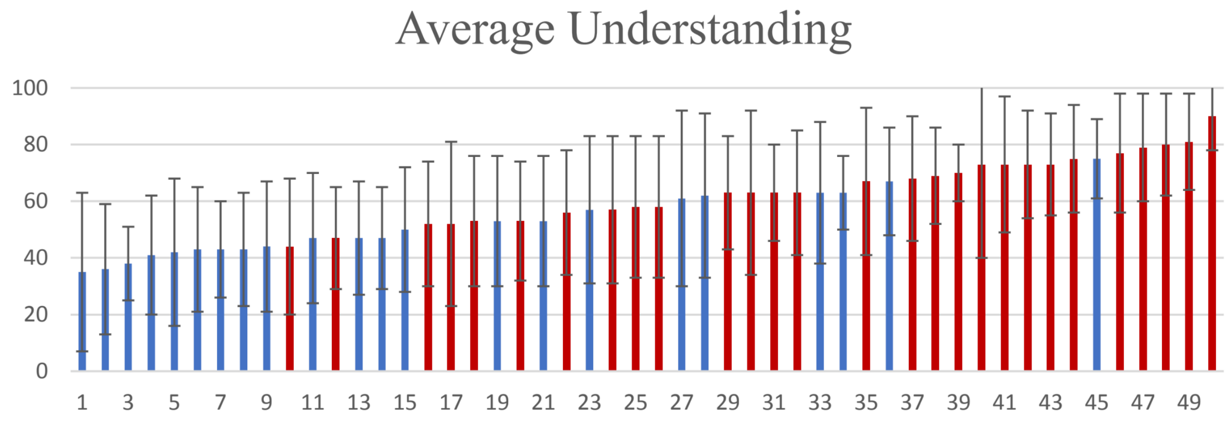

Figure 1: Round two concepts ranked by understanding

sion in the inventory. The understanding value estimates what percentage of students understand the concept, and thus a higher value of understanding means that a concept is less suited for the inventory. A positive correlation between the two indicates that the more difficult concepts are less likely to be considered important. The job of selecting concepts for inclusion in the inventory would be easier if the important concepts were also difficult; there is little value in testing mastery of an unimportant concept, and there is little need to test mastery of a concept that almost everyone understands.

The second additional question was the one that asked whether the concept was appropriate at the graduate level rather than the undergraduate level. As discussed above, Round One results for the grad/undergrad question were considered unreliable because of potential confusion from the format of the question with two independent checkboxes, and Round Two's questionnaire was modified accordingly. Analysis for Round Two showed that fewer than 20 of the 50 concepts were considered by some of the panel to be graduate-only concepts. Of these concepts, none had a majority of panelists agreeing to the graduate-only level. Therefore, while it appears, unsurprisingly, that graduate-level concepts were deemed in general more difficult to understand, there was no consensus to remove any concepts from the list based solely on inappropriateness for undergraduates.

A comparison of the ranking of the concepts by importance and understanding can be illustrated graphically. Consider Figure 1, which shows the understanding level for each of the 50 concepts considered in Round Two. Lower values on the left side of the graph (as low as 35% of students who understand the concept, on the left end) are of greater interest than the higher values on the right (which reach as high as 90% understanding on the right end). At the same time, the concepts for which the average importance is at least three on a four-point scale have a red bar rather than a blue one, and these are concentrated on the right side of the graph. The eight most difficult to understand concepts as ranked by the panel do not pass a nominal threshold for importance, while thirteen of the fourteen concepts considered easiest to understand are also those that are considered more important.

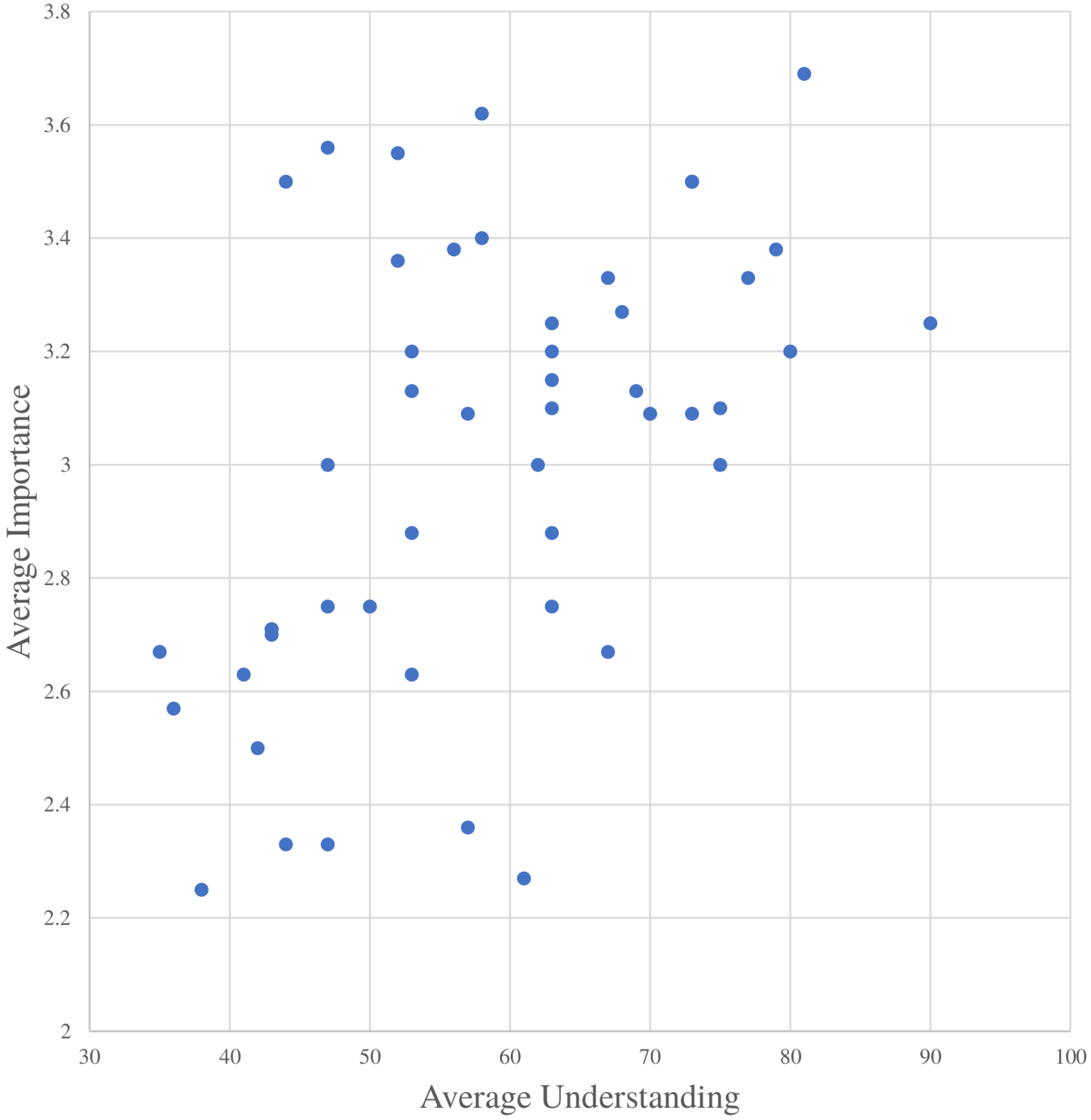

Figure 2: Scatter plot of round two concepts by importance and understanding

3.3 Selection of concepts as multi-criteria decision

Since the Keep question and the graduate/undergraduate question are not used in the final selection, the reduction of the list of concepts is a decision problem with two criteria. Figure 2 shows a scatter plot of the fifty concepts ranked in Round Two with understanding values on the horizontal axis and importance values on the vertical axis. The most promising concepts will be to the left and near the top of the graph. The reader should be able to see a pattern, however, that the points are spread out from the bottom left (low understanding, but also low importance) to the top right (high importance, but also high understanding). A concept that is more desirable than another considering both criteria, understanding and importance, is *dominant* in the Pareto sense and will appear above and to the left of the dominated concept in the graph. Even if we assume that the desirability of a concept for the concept inventory is completely captured by its importance and understanding scores – which would assume that the measurement of importance and understanding by surveys is precise, probably not a realistic assumption – there is no "natural" ranking of the 50 concepts considered in Round Two.

Two methods for imposing a ranking structure were considered. Both techniques partition the 50 concepts into precedence groups. The first method used Pareto dominance to sort the concepts into groups. The second method imposed successive thresholds on both importance and understanding.

Sorting by Pareto dominance is formulaic and requires no judgment calls. There is no one concept that dominates all the others, but there is a set of seven concepts whose members dominate the other 43 without any dominance relations among themselves. These range from an importance of 3.69 coupled with an understanding of 81 to an importance of 2.67 and an understanding of 35. We refer to these concepts as having *dominance level 1*. If we then remove this set from consideration and assemble the set of the remaining concepts that are again mutually non-dominant but together dominate all others, we have the set of five concepts at *dominance level 2*, and so on. Altogether there are ten dominance levels ranging in size from three concepts to twelve. The scatter plot of concepts is shown again in Figure 3 with the dominance levels indicated by color.

The second approach is to set arbitrary threshold values for importance and understanding separately, and to group together those concepts that exceed both thresholds. There are only two concepts that have an importance of at least 3.5 and an understanding of at most 50; there are twenty concepts with an importance of at least 3.0 and an understanding of at most 70. Grouping concepts using this method tends to rank the extremes lower: those concepts that were in high Pareto dominance levels because they had only either very low understanding or very high importance are excluded from the higher threshold groups. The scatter plot of concepts is shown once again in Figure 4 with the seven threshold levels referred to as *bands* and indicated by color.

3.4 Comments from Delphi experts

Generating the list of concepts is not only a matter of examining the average ratings provided by the panel. The experts also provided comments on most of the concepts. Round Two allowed experts to critique and expand on comments from Round One. Here, some concepts began to diverge in having more non-trivial comments, whether in quantity or quality. This may have indicated

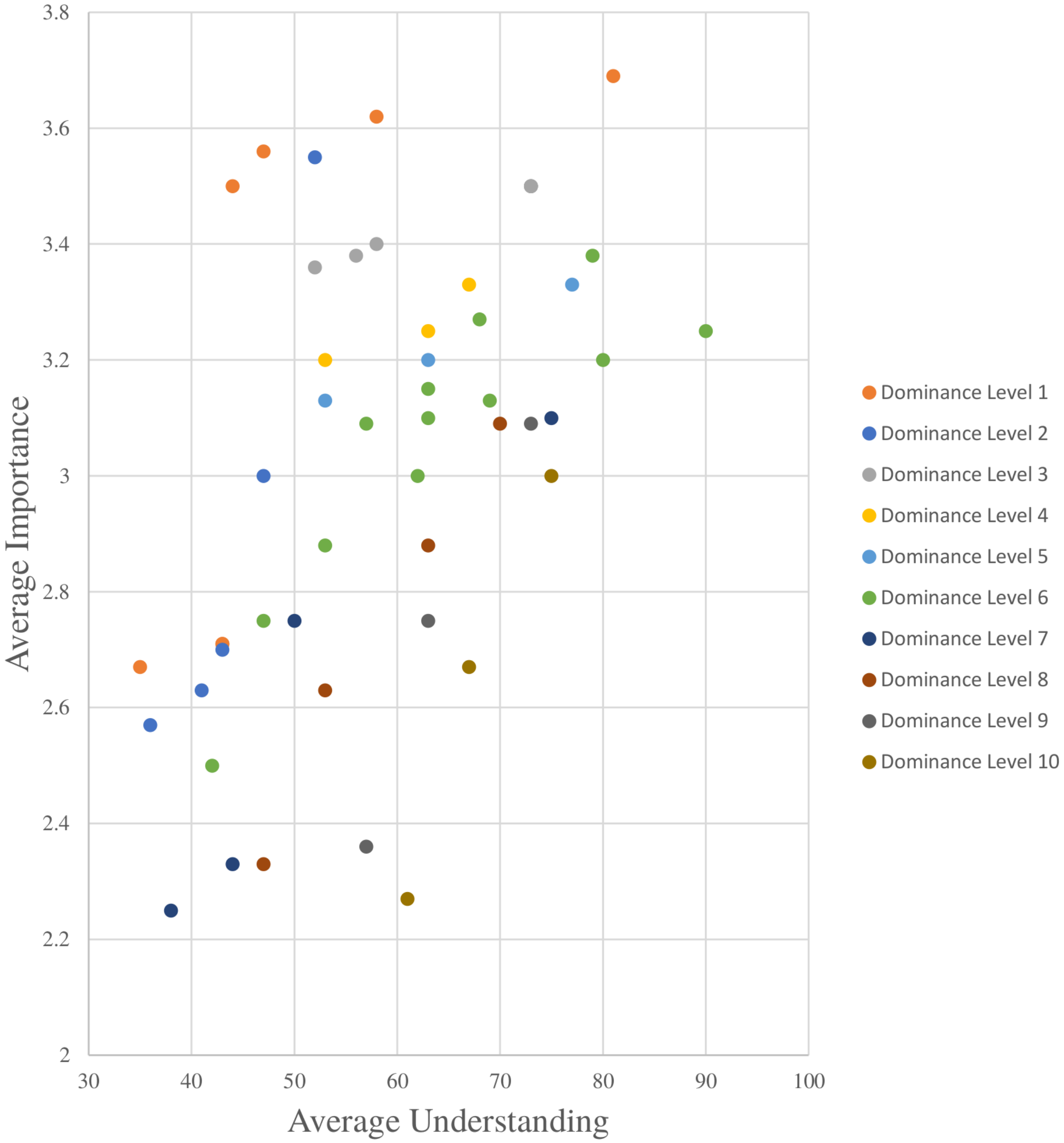

Figure 3: Scatter plot of round two concepts with dominance levels

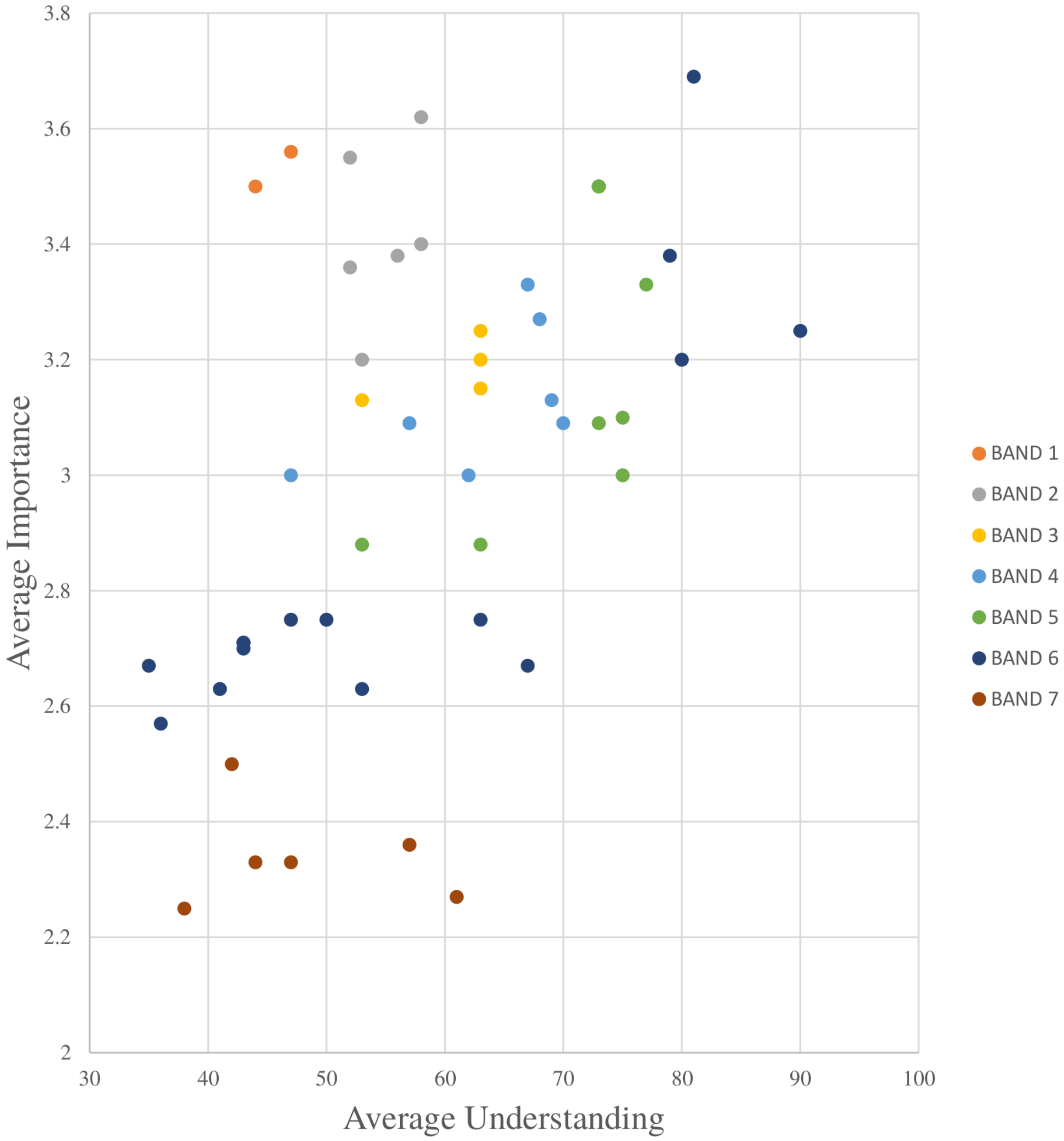

Figure 4: Scatter plot of round two concepts with arbitrary thresholds

concepts with confusing language selection or debating the overall relevance to undergraduate education. As a result of these comments, some concepts were combined. One concept had the wording changed, when several panelists objected to the use of the word “signal” to describe a flow. The panelists much preferred the word “information”. This may be a simple evolution of language, with “information” carrying a technical connotation today that it did not always carry when Pahl and Beitz [1] was translated, or it may be something more significant.

Comments from the panel will also be used at the next stage of the process when multiple-choice questions are developed and tested. The discussions between Round One and Two will be used as additional input in the next stage of concept inventory devleopment, which is the creation of multiple-choice question to attempt to identify particular misunderstandings of different concepts.

4 List of concepts for further consideration

The set of all fifty concepts from Round Two of the Delphi process is listed here. The “Threshold?” column indicates those with importance at least 3.0 and understanding at most 70:

Concept	Importance	Understanding	Dominance Level	Threshold?
Functions can be achieved by different means.	3.69	81	1	No
Functions differ from embodiment.	3.62	58	1	Yes
A function may have one of several possible operating states, or be inactive, under different conditions.	3.56	47	1	Yes
Functions can be intended or unintended; unintended functions may be unwanted.	3.5	44	1	Yes
Function models can be constructed through enumeration of flows	2.71	43	1	No
There is a distinction between overall (or "purpose") functions that refer to desired output effects and supporting (or "technical") functions that refer to the capability to execute the desired task.	2.71	43	1	No
Function models for new products may be created by three approaches: starting with the desired outputs ("backward chaining"), starting with the available inputs ("forward chaining"), or starting with central functions ("nucleation").	2.67	35	1	No
Functions are solution-neutral expressions of system operation; they express what a system does, not how the system does it.	3.55	52	2	Yes
Subfunction structure can dictate physical layout.	3	47	2	Yes
Functional descriptions cannot fully contain a designer's intent	2.7	43	2	No
Function models can be created using function-means mapping.	2.63	41	2	No
Function transformations require energy inputs.	2.57	36	2	No
Flows are energy, material, or information/signal.	3.5	73	3	No
Functions produce transformations in flows of energy, material or information/signal.	3.5	73	3	No
A function is a verb.	3.5	73	3	No
Defining functions can help identify system boundaries.	3.4	58	3	Yes
Function modeling is useful in defining/scoping the problem space (identifying new requirements and relations/dependencies between requirements).	3.38	56	3	Yes
Defining the system boundaries can help identify functions.	3.36	52	3	Yes
Functions in models are accompanied by nouns describing the flows on which they are performed.	3.33	67	4	Yes

Concept	Importance	Understanding	Dominance Level	Threshold?
A function (verb) acts on a (grammatical) object, which is transformed.	3.25	63	4	Yes
Flows are conserved across functions.	3.2	53	4	Yes
A design artifact may perform more than one function	3.33	77	5	No
High-level function models are useful for initially defining solutions.	3.2	63	5	Yes
Grouping subfunctions into modules corresponds to function sharing by components.	3.13	53	5	Yes
Functions may have multiple input and output flows	3.38	79	6	No
The aggregation of functions performed to transform a set of system inputs to a set of desired system outputs is known as a functional model.	3.27	68	6	Yes
Function models can be created by decomposition/dissection of existing products/systems.	3.25	90	6	No
A function's input can be a preceding function's output.	3.2	80	6	No
Function models for new products can be created from system requirements or customer needs statements.	3.15	63	6	Yes
Input and output flows match at all levels of a model's hierarchy.	3.13	69	6	Yes
Functions conserve material/energy at all model levels.	3.1	63	6	Yes
The allocation of functions to design artifacts determines a system's architecture.	3.09	57	6	Yes
Functions may be expressed at multiple levels of abstraction.	3	62	6	Yes
Flows are the connections of functions.	2.88	53	6	No
Functions can be temporal.	2.75	47	6	No
There are different functional modeling approaches which serve different purposes	2.5	42	6	No
A function structure can graphically represent a functional model using subfunctions and flows.	3.1	75	7	No
Function transformations result in energy losses.	2.75	50	7	No
Low-level function models are useful for value engineering and redesign.	2.33	44	7	No
Level of detail in a functional model constrains detail in a design.	2.25	38	7	No
Some functions can be decomposed into more elementary subfunctions.	3.09	70	8	Yes
Function modeling evaluates and enforces conservation principles of energy and material.	2.88	63	8	No
Function models can be created using Black Box Modeling.	2.63	53	8	No
Function modeling can be used to identify redundant flows (transformation paths).	2.33	47	8	No
Individual functions can be chained together.	3.09	73	9	No

Concept	Importance	Understanding	Dominance Level	Threshold?
Functions can express purpose, requirements, behavior, actions, or activities.	2.75	63	9	No
Standard taxonomies have been developed for articulating functions in engineering design.	2.36	57	9	No
Flows are conserved across functions.	3.2	53	4	Yes
A design artifact may perform more than one function	3.33	77	5	No
A functional model describes the process of subfunctions achieving an overall task.	3	75	10	No
Functions have purposes and benefits	2.67	67	10	No
Function models can be created using the Functional Basis.	2.27	61	10	No

5 Conclusion

Three rounds of a Delphi process to develop a concept inventory for functional reasoning have been completed. The process was modeled on the one used to create a concept inventory for thermodynamics [22]. The process to make the functional reasoning inventory has appeared more challenging than was reported by the authors of the thermodynamics instrument, beginning with Round Zero, the step in which the panel of experts is asked to provide candidate concepts for inclusion in the inventory. Only a handful of experts volunteered possible concepts, and the list used in the surveys was completed by a review of the literature. Note that the scholarly literature in functional reasoning is considerably broader than what is typically taught in undergraduate engineering classes. One Delphi panelist suggested that their published works could provide a good list of functional concepts.

Participation in the evaluation rounds of the Delphi process was much higher than in Round Zero, which nearly all of the panelists participating, and there is now a list of concepts that represents a broad consensus of the panel of experts. Once again, however, the development of the process diverged from the thermodynamics inventory development. The recommendation is to consider for inclusion in the inventory those concepts which rank high in importance and low in understanding. Unfortunately, for the functional reasoning inventory importance and understanding were largely correlated. For example, the seven concepts that had the lowest understanding ratings also had low ratings on importance. On the other hand, four of the six deemed most important did meet the threshold for understanding to be advanced.

The importance and understanding rankings can be combined in several different ways in order to determine which concepts should be moved forward in the process. Both Pareto dominance and threshold banding were considered as selection criteria. In the end, a set of 20 concepts were selected that achieved both an importance level of 3.0 on a 4-point scale and an understanding level no greater than 70%. Multiple-choice questions are currently under development for this set of candidate concepts, and will be tested with student groups in order to validate them for inclusion in the concept inventory.

6 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1504851. This support is gratefully acknowledged. The anonymous reviewers' comments are much appreciated, and the author regrets not having been able to provide more substantive responses to some of those comments.

References

[1] G. Pahl, W. Beitz, J. Feldhusen, and K.-H. Grote, *Engineering Design: A Systematic Approach*, 3rd ed. London: Springer-Verlag, 2007.

[2] C. L. Dym, P. L. with Elizabeth J. Owen, and R. E. Spjut, *Engineering Design: A Project-Based Introduction*, 3rd ed. New York: John Wiley, 2009.

[3] R. L. Nagel and M. R. Bohm, "On teaching functionality and functional modeling in an engineering curriculum," in *ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2011, pp. 625–636.

[4] M. S. Erden, H. Komoto, T. J. van Beek, V. D'Amelio, E. Echavarria, and T. Tomiyama, "A review of function modeling: Approaches and applications," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 22, no. 02, pp. 147–169, 2008.

[5] A. Chakrabarti and T. P. Bligh, "A scheme for functional reasoning in conceptual design," *Design Studies*, vol. 22, no. 6, pp. 493–517, 2001.

[6] A. Chakrabarti, P. Sarkar, B. Leelavathamma, and B. Nataraju, "A functional representation for aiding biomimetic and artificial inspiration of new ideas," *AIE EDAM*, vol. 19, no. 02, pp. 113–132, 2005.

[7] Y. Umeda, T. Tomiyama, and H. Yoshikawa, "Fbs modeling: modeling scheme of function for conceptual design," in *Proceedings of the 9th international workshop on qualitative reasoning*, 1995, pp. 271–278.

[8] Y. Umeda, M. Ishii, M. Yoshioka, Y. Shimomura, and T. Tomiyama, "Supporting conceptual design based on the function-behavior-state modeler," *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing*, vol. 10, no. 04, pp. 275–288, 1996.

[9] M. Yoshioka, Y. Umeda, H. Takeda, Y. Shimomura, Y. Nomaguchi, and T. Tomiyama, "Physical concept ontology for the knowledge intensive engineering framework," *Advanced Engineering Informatics*, vol. 18, no. 2, pp. 95–113, 2004.

[10] K. D. Forbus, “Qualitative process theory,” *Artificial intelligence*, vol. 24, no. 1, pp. 85–168, 1984.

[11] Y.-M. Deng, “Function and behavior representation in conceptual mechanical design,” *AI EDAM*, vol. 16, no. 05, pp. 343–362, 2002.

[12] Y. Umeda and T. Tomiyama, “Functional reasoning in design,” *IEEE expert*, vol. 12, no. 2, pp. 42–48, 1997.

[13] Foundation Coalition, “Concept Inventory Assessment Instruments,” http://www.foundationcoalition.org/home/keycomponents/concept/index.html, 2009, accessed October 2, 2014.

[14] ciHUB, “cihub - resources: Concept inventories,” http://cihub.org/resources/conceptinventories/?tag=engineering, 2011, accessed October 21, 2014.

[15] J. W. Pellegrino, L. V. DiBello, and S. P. Brophy, “The science and design of assessment in engineering education,” in *Cambridge Handbook of Engineering Education Research*, A. Johri and B. Olds, Eds. Cambridge, England: Cambridge University Press, 2014, pp. 571–598.

[16] D. Hestenes, M. Wells, and G. Swackhamer, “Force Concept Inventory,” *The Physics Teacher*, vol. 30, no. 3, pp. 141–151, 1992.

[17] J. Epstein, “Development and validation of the Calculus Concept Inventory,” in *Proceedings of the Ninth International Conference on Mathematics Education in a Global Community*, Charlotte, NC, 2007.

[18] J. W. Pellegrino, N. Chudowsky, and R. Glaser, Eds., *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press, 2001.

[19] K. Midkiff, T. Litzinger, and D. Evans, “Development of engineering thermodynamics concept inventory instrument,” in *Frontiers in Education Conference*, Reno, Nevada, 2001, 1013 October 2001.

[20] P. Steif and J. Dantzler, “A statics concept inventory: Development and psychometric analysis,” *Journal of Engineering Education*, vol. 94, no. 4, pp. 363–371, 2005.

[21] S. Krause, J. Decker, and R. Griffin, “Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses,” in *33rd Annual Frontiers in Education (FIE’03)*, vol. 1, 2003, pp. T3D–11.

[22] R. A. Streveler, R. L. Miller, A. I. Santiago-Roman, M. A. Nelson, M. R. Geist, and B. M. Olds, “Rigorous methodology for concept inventory development: Using the ‘assessment triangle’ to develop and test the Thermal and Transport Science Concept Inventory (TTCI),” *International Journal of Engineering Education*, vol. 27, no. 5, pp. 968–984, 2011.

[23] M. J. Clayton, “Delphi: a technique to harness expert opinion for critical decisionmaking tasks in education,” *Educational Psychology*, vol. 17, no. 4, pp. 373–386, 1997. [Online]. Available: http://dx.doi.org/10.1080/0144341970170401