AC 2012-3947: DEVELOPMENT AND DEPLOYMENT OF A RUBRIC BASED ON FINK'S COGNITIVE DIMENSIONS IN A FLUID MECHANICS AND HEAT TRANSFER CLASS WITH POTENTIAL APPLICATIONS IN A VARIETY OF ENGINEERING CLASSES

Mr. Baba Abdul, Washington State University

Baba Abdul obtained an M.Sc. in chemical engineering from the Ahmadu Bello University (ABU), Zaria, Nigeria, in 2005, and has had some work experience in the chemical process industries, mainly crude oil processing, refining, and solids processing. He is currently working on a Ph.D. that includes elements of fluid mechanics in small helicosymmetric channels and engineering education.

David B. Thiessen, Washington State University
Prof. Bernard J. Van Wie, Washington State University

Bernard Van Wie has been teaching for 30 years, first as a graduate student at the University of Oklahoma, and then as a professor at Washington State University. Over the past 15 years, he has devoted himself to developing novel teaching approaches that include components of cooperative/collaborative, hands-on, active, and problem/project-based learning (CHAPL) environments.

Dr. Gary Robert Brown, Portland State University
Dr. Olusola O. Adesope, Washington State University, Pullman

Olusola O. Adesope is an Assistant Professor of educational psychology at Washington State University, Pullman. His research is at the intersection of educational psychology, learning sciences, and instructional design and technology. His recent research focuses on the cognitive and pedagogical underpinnings of learning with computer-based multimedia resources, knowledge representation through interactive concept maps, meta-analysis of empirical research, and investigation of instructional principles and assessments for engineering designs.

Development and Deployment of a Rubric Based on Fink's Cognitive Dimensions in a Fluid Mechanics and Heat Transfer Class with Potential Applications in a Variety of Engineering Classes

Abstract

There is a need for better and more objective assessments of student cognition especially at the engineering college level. This is important to capture essential intellectual abilities that may be missed through conventional testing, produce assessments that are far more descriptive of student cognition than a single holistic grade, link learning outcomes and professional competencies, enable multiple evaluators to apply the same criteria to judge the same work and design better formative interventions. This paper reports on the development and subsequent deployment of a rubric based on Fink's cognitive dimensions of Foundational Knowledge (F), Application (A) and Integration (I) in a Fluid Mechanics and Heat Transfer (FMHT) class though it has potential broader multi-disciplinary applicability. Professors from Chemical Engineering, Mechanical Engineering and Education alongside graduate students from Chemical Engineering were involved in the development of the rubric. Definitions of what exactly constitute F, A and I in the subject domain were negotiated, scales indicating levels of performance were agreed upon and a minimum competency "anchor" line was drawn. Deployment of the rubric was done following a Convergent Participation Model (CPM) to be described in the body of the paper. Three main types of assessment artifacts namely traditional text-book problems, professor-crafted worksheets and final examination problems were rated for the aforementioned cognitive dimensions. A portion (25%) of the anonymous artifacts were selected and rated by 2 different panels consisting of professors and graduates students involved in the rubric development, or otherwise trained in its use. Our premise was that we would see consistent ratings for the F, A, and I dimensions and that ratings for the worksheets would be identical to those for traditional textbook problems. However, the three dimensions were found to produce different ratings, depending on the depth and complexity of the assignment, and the three types of assessment artifacts were found to differ significantly in rubric outcomes. These and other interesting findings are discussed with a view to designing better learning experiences and assessments.

Keywords: Cognition, Rubric, Convergent participation model,

I. Introduction

Performance assessment is ubiquitous in human experience and includes all tasks that involve making judgments (as objective as possible) on a wide range of human endeavors. It is especially crucial in the educational enterprise because education prepares humans for all other spheres of human endeavor. Educational assessment is used in many important situations. These include making judgments about the competency level of a learner, the form and timing of learning interventions and enrollments/hiring/admissions.

As stakeholders, especially the engineering education accreditation body and employers of labor demand better quality engineering graduates[1, 2], engineering educators are tasked with coming up with more effective pedagogies[2-3] and associated assessment strategies to meet this demand. The importance of performance assessment cannot be overemphasized as it provides, amongst other things, a tool for helping learners to more successfully navigate the trajectory of “hard” cognition[2] and “professional” or “soft” skills[2].

Various assessment tools have been developed, tested and deployed in attempts to capture learner cognition [4-5, 6-7]. Concept inventories[4, 5, 8] which are usually multiple-choice questions are often limited to capturing learner conceptual knowledge in the domain of interest. However, Zhao[9] has highlighted pitfalls in the design and interpretation of multiple-choice artifacts and suggests ways to mitigate the effects of guesswork. Another cognitive assessment tool is personal communication especially in form of interviews and listening to learner group discussions[10-11]. Marra et al[11] used Perry’s model[12] to classify interview results according to 9 levels of cognitive attainment. While interviews may yield more information than say, multiple-choice tests, they are resource-intensive and hence less efficient[13]. Rubric scoring of assessment artifacts such as reports[14], essays[15, 16] and quantitative/qualitative problems[16-17] are becoming more popular in engineering education because they are more informative and more objective than a single assessor-assigned score. However, care needs to be exercised to ensure that the rubric conforms to the theory of how cognition is developed in the domain of interest[18] and that it is adequately explicit and sensitive enough to discriminate between artifacts of different quality[19-20].

Rubrics have been designed and used to capture cognition in the domain of physical transport phenomena (fluid flow, mass and heat transport). Yadav and co-workers[21] used a 4-point rubric to grade a thermal and fluid systems modeling class. However, a detailed rubric was not provided to enable wider use, and vague descriptors like no understanding, average, good and excellent understanding were used for the points on this scale. Even though the rubric was not detailed enough to enable external perusal, the authors reported respective mean scores of 2.17 and 2.08 for the thermal and hydraulics design dimensions of the test. Another noteworthy rubric used in fluid and thermal cognition assessment is the critical thinking rubric adapted by Golter et al[6-7]. This is a 5-point, 9-dimension rubric which was found very useful in grading student projects in a project-based class[22]. We acknowledge that a subset of this rubric could also be used to grade short homework assignments, but posit that the descriptors are not explicit enough and do not adequately depict how cognition and expertise is developed in the domain. We therefore propose a concise yet detailed 5-point, 3-dimension Fink’s[23] Foundational Knowledge, Application and Integration (FAI) rubric which we believe adequately mirrors cognition in the macroscopic analysis of fluid and thermal systems. We believe this rubric is concise, more explicit and relatively easy to use and can be employed in the rating of all the assessment artifacts by a trained rater. The current version of this rubric is appended at the end of this paper. Our hypotheses are: 1. that we would find consistent FAI ratings for all assessment artifacts (text book problems, worksheets and examinations) that would correlate with ratings obtained from traditional professor-crafted scoring keys, 2. the rubric scores would be more descriptive and hence more representative of learner cognitive trajectory. This we believe would be more useful for formative purposes.

II Theoretical Considerations

In this section we discuss some of the theoretical underpinnings of this work. These include the theories of assessments and rubric design.

Purposes of assessment

The major purposes of educational assessment like any other type of assessment are twofold: to 1. provide formative and/or 2. provide summative reviews[18]. Formative assessment gives information that is used to correct learning deficits while summative assessment gives information about achievement at the end of a learning or appraisal exercise. Crouch and Mazur used formative assessment to enhance a peer-instruction protocol at Harvard and reported significant normalized gains on the Force Concept Inventory (FCI)[8] over an eight-year period (p 131 of reference[24]) It is widely believed that a thorough and comprehensive assessment program can assure the quality of practitioners and practice in a domain and that this exercise is better left to practitioners in the domain rather than experts from social sciences[25]. Transcripts of courses passed are no longer considered adequate to support claims of what students know. Mapping outcomes to courses and relating them to traditional assessments such as tests and quizzes gives more direct evidence of student learning. ABET EC-2000 criterion 3 defines these program level outcomes in the larger domain of engineering which are necessarily supported by course level outcomes[25-26].

Regardless of the *purpose* of *assessment*, it is imperative that the two be in alignment in the specific domain [7, 18, 24]. Assessment and associated tools must align with the subject domain, learning objectives and learning theories and must also be unbiased for validity and fairness (p 39 of [18]).

In most educational settings the following learning objectives may be assessed: knowledge, skills and attitudes (or dispositions) [13, 27]. Stiggins [13] shows an example of mapping learning objectives and assessment methods. The methods include selected response, essay, performance and personal communication. More than one of these methods can be used to assess each outcome based on alignment. In this study a mixed quantitative/qualitative essay was used to assure a quality learning experience.

Inferential nature of assessment.

Because we cannot directly measure what a student knows like we would his/her weight, we can only infer an examinee's competence from observation of an assigned task [18]. The assessor reasons from observable evidence relevant to the domain, learning theory and objectives using relevant and available tools. This process can be represented in what is termed assessment triangle. Shown in Fig. 3 it depicts the connection between cognition, observation and interpretation.

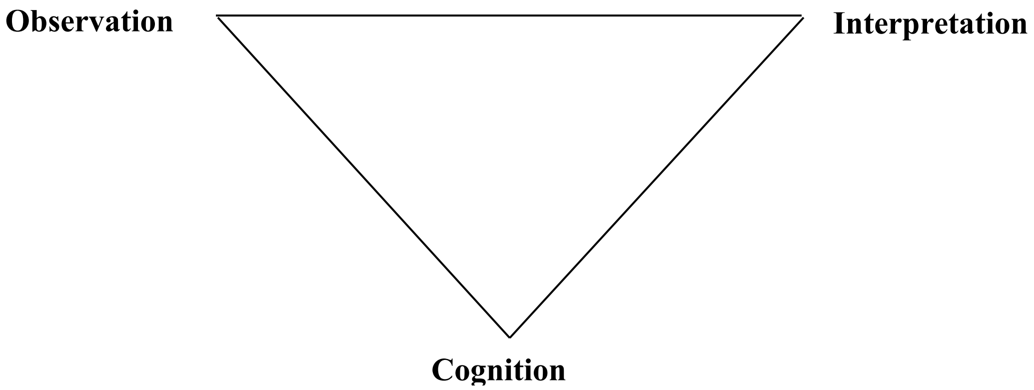


Figure 1: Triangle showing elements important in making judgments from evidence of assessment.

Cognition: This refers to idealized ways in which learners typically represent knowledge and develop expertise in a given domain for a given learning objective. Streveler and co-workers [4] posit that conceptual knowledge is invaluable on the trajectory to expertise in the engineering sciences of which thermal and transport science is one.

Observation: This is an assigned task that will elicit behaviors which show the important knowledge, skills or attitudes that the assessor is testing for. This should be carefully designed to give evidence linked to the cognitive model and support the inferences and decisions based on it.

Interpretation: This includes all the tools and methods used to reason from fallible observations. In other words, the observations derived from a set of assessment tasks is used to rate the degree of attainment of the learning outcomes.

For effective assessment, all the three vertices of the triangle must be synchronized. The cognitive theory of how expertise is represented in that domain gives clues as to kind of activities that will elicit signs of that competence and how to interpret the signs. Also, the evidence gathered should be of the type that the assessor has the tools to interpret as it relates to the targeted learning outcome. An instructive quote in Turns and co-workers paper section aptly titled ‘Opportunities for future research’[28], succinctly captures the importance of this triangle:

> “One priority for research is the development of cognitive models of learning for areas of the school curriculum. …researchers have developed sophisticated models of student cognition in various areas of the curriculum, such as algebra and physics. However, an understanding of how people learn remains limited for many other areas. Moreover, even in subject domains for which characteristics of expertise have been identified, a detailed understanding of patterns of growth that would enable one to identify landmarks on the way to competence is often lacking. Such landmarks are essential for effective assessment design and implementation”

Authenticity of assessment:

Assessment is authentic (real, genuine, applied or authoritative) when student performance is directly examined on worthy intellectual tasks [29, 30]. Engaging teaching methods and assessments that emphasize conceptual knowledge rather than rote skills and recall have been reported to promote deeper learning [31]. According to Lauren Resnick [30] “What you assess is what you get; if you don't test it you won't get it. To improve student performance we must recognize that essential intellectual abilities are falling through the cracks of conventional testing”. In other words, if assignments and assessment are not designed to adequately capture learner competencies, important formative insights may be lost. For instance, if knowledge, application and integration are not rigorously examined, important information about how learners relate these dimensions will not be grasped. In this work therefore, we have attempted to design assessment accordingly with a view to using the results for curricular and pedagogical reform. Table 1 presents a concise comparison of authentic and traditional assessments.

Table 1: Comparison of authentic and traditional assessments.

Authentic Assessment	**Traditional Assessment**
Rigorous	Simplistic
Ill-structured	More like drills
Validity depends on ‘real-world’ test of ability	Validity is mainly dependent on statistics
Achieves validity and reliability by emphasizing and standardizing the appropriate criteria for scoring such (varied) products	Standardizes objective "items" and, hence, (one) right answer for each
More challenging, answers are creative	One-answer questions
Expensive	Cheaper
Mostly based on human judgment	Machines can score this
Makes students thinking visible to themselves and others	Often a black box

Interconnections of assessment with content and pedagogy

Assessment is the means used to measure the *outcomes* of a *pedagogy* employed to deliver the contents in a particular discipline or domain. The three should of a necessity be aligned such that they support each other for learning efficiency [32]. Furthermore, decisions on instruction and assessment should be based on the current best model of learning in the domain.

Significant learning and Fink's taxonomy

Fink defined significant learning as learning that would "produce a lasting change in terms of the learner's life" and proposed a non-hierarchical, relational and interactive taxonomy [23] that he believed could succeed the popular though hierarchical Bloom's taxonomy [33]. This taxonomy transcends the classical Bloom's cognitive taxonomy in two main ways 1) includes other objectives like learning about learning, ethics, team skills and character, which do not easily emerge from Bloom's taxonomy and 2) highlights the non-hierarchical, relational and synergistic character of the learning objectives (i.e. gains in one dimension reinforces some other dimension).

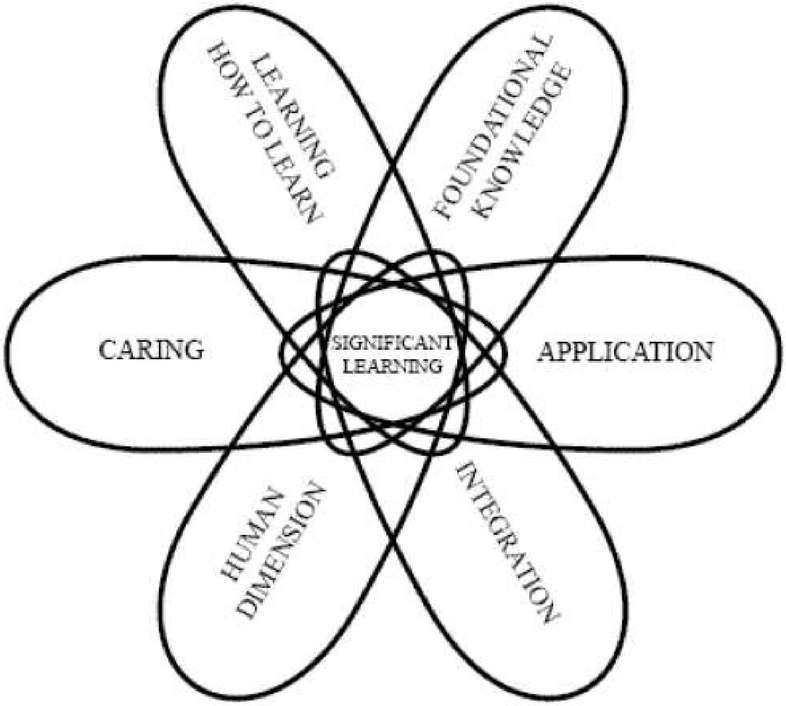


Fig. 2. The relation between the six Fink dimensions.

Figure 2 illustrates the relational character of the learning objectives in Fink's taxonomy. These learning objectives are: foundational knowledge, application, integration, human dimension, caring, and learning how to learn. A brief description of each is given below:

Foundational knowledge- Foundational knowledge in FMHT can lead to better conceptualization and lead to other types of learning. In the course at hand students should understand and remember basic ideas like laminar/turbulent flow, dimensionless numbers, their significance and limits, correlations and their significance.

Application- Involves learning new skills (critical, creative and practical) by engaging in some kind of action. It enables other kinds of learning to be useful. Students solve FMHT problems by applying basic knowledge. In a course such as ours by solving problems they develop the aforementioned skills. For instance they get a feel for what is practical and what is unattainable. They also get a feel for how real systems (e.g. DLMs) differ from idealized systems as represented by theoretical equations.

Integration- It is important for students to have the ability to connect ideas from different sub-domains because rarely are real-world engineering problems isolated. For instance, since the chemical engineering systems in this study are mostly flow systems, situations of simultaneous momentum and heat transfer are the norm rather than the exception, and so knowledge of the nature of fluid flow is necessary to predict nature and quantity of heat transfer.

Human dimension- Learning about the social dynamics involved in a learning community can help the learners learn how to interact effectively. Engineering practice is essentially team-based hence this learning objective can better prepare students for professional life.

Caring-Caring about something gives the learner more energy to learn more about it. For instance, learning about the societal impact of fluid mechanics and heat transfer can motivate students to learn more about this important field.

Learning how to learn-When students learn about the process of learning itself (for instance learning about scientific method or how to be a self-directed learner), it helps them to be more effective lifelong learners. We believe that the kind of assignments in the worksheets used in this study helps to promote self direction in the learning process.

Design of scoring rubric

According to the Teaching, Learning and Technology (TLT) website [20], "**a *rubric* is an explicit set of criteria used for assessing a particular type of work or performance.** A rubric usually also includes levels of potential achievement for each criterion, and sometimes also includes work or performance samples that typify each of those levels. Levels of achievement are often given numerical scores". Rubrics provide an effective, efficient equitable assessment that can also be understood and applied by the learner as well as the teacher-assessor [19] The following are some reasons for using rubrics [20]:

1. To produce assessments that are far more descriptive than a single, holistic grade or judgment can be.

2. To provide a richer and more multidimensional description of the reasons for assigning a numerical score to a piece of work.

3. To enable multiple judges to apply the same criteria to assessing work. For example, student work can be assessed by faculty, by other students and by working professionals in the discipline.

4. To enable comparison of works across settings. For example, imagine an academic department trying to develop skills A-G among their students. One first year course focuses on teaching goals A, B, and D, while another first year course teaches A, C, and E. One second year course is trying to deepen skill B while introducing skill E. And so on. If faculty uses the same rubrics and then pool data, the department can monitor student progress as they work toward graduation. It's a far more informative way to assess student progress and guide changes in the curriculum than to monitor student GPAs: faculty can see which skills are developing as hoped, and where there are systemic problems in teaching and learning.

   Allen and Knight [19] propose that rubrics should be collaboratively developed, if at all possible, by academics and practitioners in the field of interest to ensure validity, reliability, and fairness. They propose an eight-step guide for doing this. These steps are summarized below:

   1. Develop learning objectives for course
   2. Link learning outcomes and professional competencies
   3. Develop rubric
   4. Longitudinally test student learning as measured by rubric, establish baseline, and refine using academic and professional input.
   5. Identify problems with sub-optimal performance.
   6. Improve construct validity of rubric (determine weights for dimensions)
   7. Determine ability of rubric to distinguish student performances (e.g. poor average or excellent)
   8. Establish inter-rater reliability and further validate rubric

III. Methodology

Rubric development

In this study an 11-scale rubric (5 points with half-point divisions, zero inclusive) used to rate the assignments from the lecture and CHAPL (Cooperative Hands-on Active Problem/Project-based Learning) topics was constructed by interpreting the relevant constructs in Fink's taxonomy [23] to suit the FMHT domain. The 3 cognitive dimensions: knowledge, application and integration were defined in the FMHT domain and levels of performance on each were described in the rubric (see appendix for rubric details). For purpose of anchoring, a grade of B minus (B-) assigned a numerical value of 3 points out of 5 was chosen as the competency line at which significant learning is deemed to commence.

Rating of Assignments

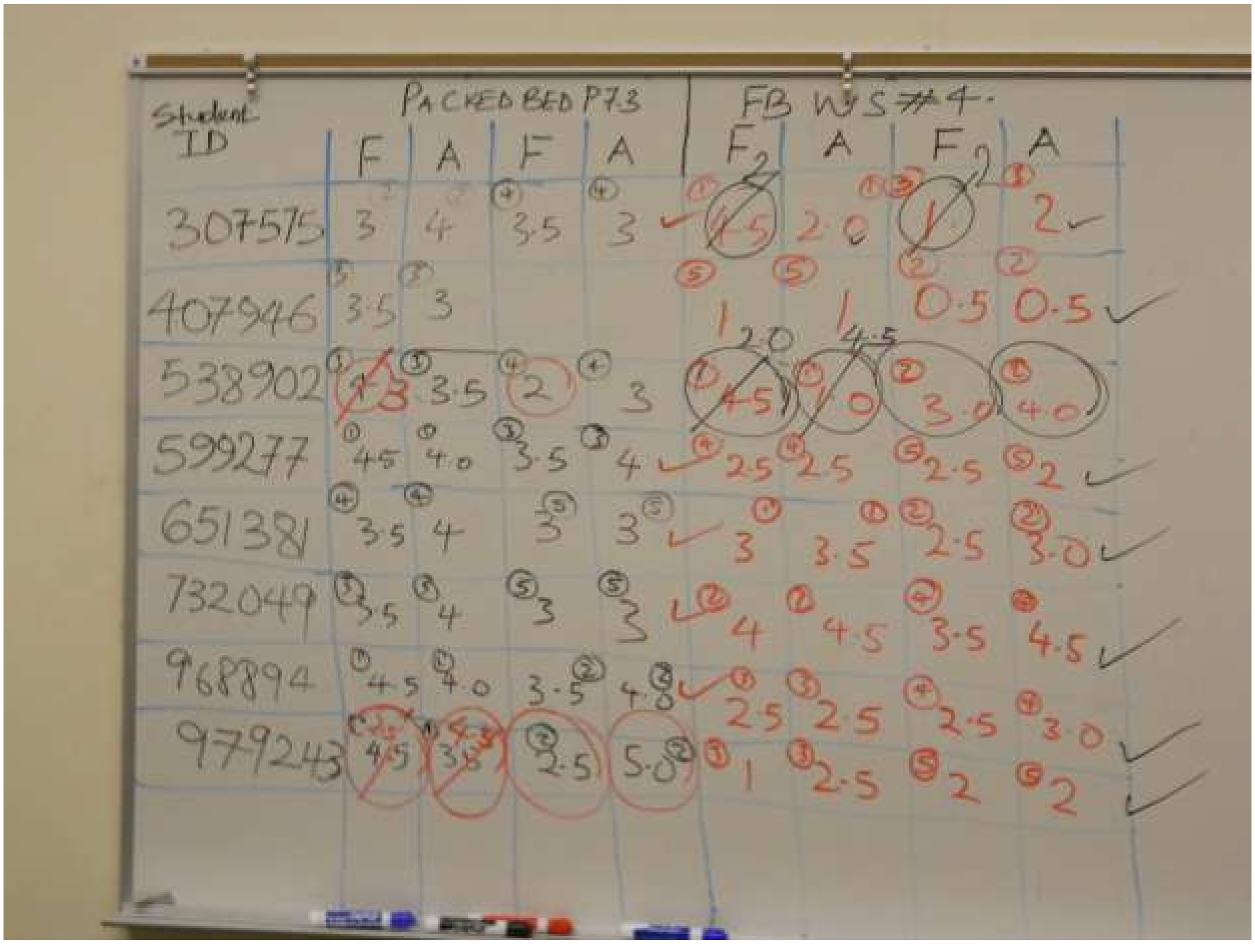
Rating, scoring or assessing a learner's product is not a trivial undertaking because of the premium placed on validity, reliability and fairness by the educational community. Therefore, for a test to be acceptable, it must be stable, consistent and repeatable. Inter-rater reliability, or the degree of agreement between two or more raters of a particular product, is believed to be necessary to both verify coherence in a quantitative study, and verify and solidify qualitative research findings [34].

In this study, the first step we took in the establishment of an acceptable inter-rater reliability (≥70% is the norm set by the Educational Testing Service, ETS) was to convene "norming" sessions. For these sessions, 2 panels were employed based on intellectual pedigree, familiarity with content knowledge and availability. The first panel consisted of 4 chemical engineering professors and 1 chemical engineering TA while the second consisted of 2 mechanical engineering professors and 3 chemical engineering PhD students. The panels rated 25% of student assignments anonymously (text book assignments and worksheets) using the rubric described in the previous subsection. Anonymity was assured in compliance with FERPA (Family Education Rights and Privacy Act) rules by generating random numbers to represent students using Microsoft Excel. These numbers were printed on each artifact assigned to a rater. During the rating, some disagreements on the structure of and the language of the rubric were resolved and led to the modification of the rubric. For instance, 2 of the raters had to negotiate what to rate as foundational knowledge and what as application. Each assignment was scored by 2 randomly assigned raters followed by a discussion moderated by a different chemical engineering PhD student skilled in such evaluation protocols, with a view to resolving any differences within at most a 1.0 margin. While the moderator urged them to do their utmost in resolving differences based on domain-specific cognition, he also cautioned against undue coercion or giving in for the sole purpose of conforming rather than personal conviction. This is akin to a convergent participation model used by Nesbit, Belfer and Vargo [35] . Figure 3 shows the results of a typical "norming" session on a white board for 2 assignments. The first assignment is on packed bed while the second is on fluidized bed. The scores that require discussions are circled in red ink. After resolution of differences, inter-rater reliability was estimated to be greater than 70%.

Deployment of rubric

After the training of raters, fine tuning and testing of rubric, the rubric was used to grade some the students' assignments. This grading was done by one of the raters who participated in the "norming", who was now believed to have adequate training to do so.

The average rubric scores where then compared to the average regular single digit scores from the teaching assistant (TA) for the class who did not participate in the rubric development or "norming" in any way whatsoever. The TA used an answer key that we consider to inadequately delineate the cognitive dimensions of importance in this course (namely, F, A and I). This was done because of the "black box" nature of lumped scores and their limited utility in making formative decisions.



Student ID	PACKED BED P7.3				FB WS #4.			
	F	A	F	A	F	A	F	A
307575	3	4	3.5	3	~~4.5~~ 2	2.0	~~1~~ 2	2
407946	3.5	3			1	1	0.5	0.5
538902	~~4~~ 3	3.5	2	3	~~4.5~~ 2.0	~~1.0~~ 4.5	3.0	4.0
599277	4.5	4.0	3.5	4	2.5	2.5	2.5	2
651381	3.5	4	3	3	3	3.5	2.5	3.0
732049	3.5	4	3	3	4	4.5	3.5	4.5
968894	4.5	4.0	3.5	4.8	2.5	2.5	2.5	3.0
979243	~~4.5~~	~~3.8~~	2.5	5.0	1	2.5	2	2

Figure 3: A depiction of a typical convergent participation process. The scores circled in red ink have wide disparity and will be discussed by the 2 raters responsible with inputs from other panel members.

IV Results and Discussion

In this section we will present and discuss the initial rating and also the ratings after discussions and negotiations. We will also present and discuss the results for the deployment of the rubric in rating sets of student assignments.

Table 2 depicts the rubric results for a particular assignment rated by 2 independent panels. This assignment is a purely fluid mechanics problem hence it is devoid of the I (integration) dimension as defined in the rubric. This is because the rubric restricted integration to assignments whose solution requires active integration of fluid mechanics and heat transfer principles

Table 2: Typical raw norming results showing scores by the 2 panels for the packed bed assignment. The scores in bold font represent scores that are in disagreement (difference>1)

1st panel												
	Initial						**Final**					
Student code	**R#**	**Dim.**		**R#**	**Dim**		**R#**	**Dim**		**R#**	**Dim**	
		F	**A**		**F**	**A**		**F**	**A**		**F**	**A**
307575	E	4.5	4.5	C	4.5	4	E	4.5	2	C	4.5	3
407946	B	**3**	**2**	D	**0**	**5**	B	3	2	D	2.5	2.5
538902	E	**4.5**	**3.5**	D	**1.5**	**5**	E	**4.5**	2	D	**3**	3
599277	B	3	**2.5**	D	2	**5**	B	3	**1.5**	D	2	**3**
651381	G	**3**	**5**	E	**4.5**	**2**	G	3	3	E	2	2
732049	G	**2.5**	**2.5**	D	**0**	**5**	G	2.5	2.5	D	2.5	3
968894	G	**3.5**	**3.5**	C	**5**	**5**	G	3.5	3.5	C	4	3
979243	E	3	**2**	C	4	**3.5**	E	3	2	C	4	1.5
2nd panel												
307575	J	3	4	L	3.5	3	J	3	4	L	3.5	3
407946	K	3.5	3	H	3	3	K	3.5	3	H	3	3
538902	K	**4**	3.5	L	**2**	3	K	**4**	3.5	L	**2**	3
599277	H	4.5	4	K	3.5	4	H	4.5	4	K	3.5	4
651381	L	3.5	4	M	3	3	L	3.5	4	M	3	3
732049	K	3.5	4	M	3	3	K	3.5	4	M	3	3
968894	H	4.5	4	J	3.5	4	H	4.5	4	J	3.5	4
979243	H	**4.5**	**3.5**	J	**2.5**	**5**	H	**4.5**	**3.5**	J	**2.5**	**5**

From this table we see that agreement (difference in scores of ≤ 1) starts at 37.5% and 12.5% for the F and A dimensions respectively and increases to 87.5% after discussions for both dimensions for the scores generated by the first panel. On the other hand, the second panel remains at constant agreements of 75% and 87.5% for the F and A dimensions respectively, before and after discussions. This could be attributed to the fact that the rubric had already been fine-tuned with the first panel that it was easier to get the second panel to reach agreement. A similar procedure was followed for the "norming" of the other assignments and inter-rater agreement of ≥ 70% was achieved after discussions.

Table 3 presents results of the scoring of some selected assignments. These assignments are of two types: traditional textbook problems and professor-developed worksheets. Note that the problems that do not have the I dimension are considered to be isolated fluid mechanics or heat transfer problems by rubric definition.

Table 3: FAI results for two types of assessment artifacts: 1.Worksheets and 2. Text book problems.

Rubric dimension	F		A		I		N	Traditional (key) score,%	
Topic	Mean	SD	Mean	SD	Mean	SD		Mean	SD
Worksheet problems									
Venturi meter	**3.17**	0.99	**3.58**	0.95			24	77.2	13.06
Fluidized bed	2.90	0.79	2.71	0.89			26	**92.13**	8.61
Shell and tube	2.44	0.58	2.73	0.67			28	79.28	8.14
Evaporative cooling	2.73	0.77	2.58	0.99	1.87	1.18	15	79.8	4.90
Textbook problems									
Pitot tube	**4.28**	0.91	**3.97**	1.26			24	92.14	14.91
Packed bed	**3.55**	0.82	**3.39**	0.99			26	93.19	5.16
Double pipe	**3.41**	0.77	**3.82**	0.65			28	95.82	4.56
Boiling	2.80	0.49	**3.23**	0.84	2.57	0.85	15	**95.70**	10.0

From table 2 the scores for the worksheets are observed to be consistently lower than those for the textbook problems. This is attributable to the fact that the worksheets are more complex than the typical text book problem. The worksheets were crafted by the professor to reflect how cognition is developed in the domain, whereas most text book problems have similar solved problems within the text (indeed some of them are partially or completely solved in the companion solutions manual). Thus text book problems may not be as challenging as the worksheets. See appendix B for a typical worksheet and text book problem.

A further look at table 2 shows in normal type face scores below the minimum competency score of 3 which would represent candidates for formative action. The most convenient formative action in this case was to post detailed problem solutions on the website for the students to study. The I (integration) dimension for the evaporative cooler worksheet shows the lowest mean score of 1.87 and a standard deviation of 1.18 indicating that a lot of students integrated the principles in this assignment rather poorly. Thus I1 would be a top priority candidate for formative action and so we made the solution for the I1 part more detailed than the others.

Further still, comparison of the traditional key-generated scores with the rubric scores indicate just how much of a "black box" the former is. For instance the key-generated scores for boiling and fluidized bed assignments are respectively 95.7% and 92.13% indicating a high performance while the rubric scores for these same assignments mostly fall below minimum competency. Hence, sole reliance on the key-generated scores will preclude any formative action on these topics.

Table 3 shows the rubric and key-generated scores for 3 of the 4 final examination problems.

Table 3: Final examination problems percentage scores and rubric scores

Rubric dimension	F	A	I	% scores
Question	Mean	Mean	Mean	Mean
1	2.64	2.57	Not applicable	68.02
2	2.19	2.13	2.16	53.58
3	2.45	2.57	2.02	65.79

An examination of this table indicates that the rubric scores fairly well correlate with the key-generated percentages. It also shows that the students on average are still below the competency line of 3. These findings will be used as formative feedback to redesign learning experience and assessment in future classes.

V Conclusions and recommendations for further work

A new rubric was developed and is now available based on three key cognitive parameters of Fink's Significant Learning Taxonomy, Foundational Knowledge, Application and Integration (FAI). Rubric results are presented then used to rate homework assignments and exam problems. Each rubric dimension contains 5 ranking levels in columns with that corresponding to the lowest proficiency spanning from 0 to 1.0, and the highest proficiency from 4.0 to 5.0. Half increment scores are allowable within each column. To assist the rater, general definitions for each column are provided and these are anchored in two ways: first, with letter grades from F (score of 0.0) to A score of (5.0) to allow professors to align student work within a standard grading framework for which they are familiar, and second, with a Yes/No Significant Learning delimiter at 3.0 or a grade of B-. Rating descriptors that are now available were developed through intense interaction between field expert professors that continued until all were satisfied with the definitions.

The rubric scores were designed and corroborated to be more descriptive of learner cognition than a single grade or score and hence more useful in future formative interventions. We aim to take these results into consideration when designing learning experience and assessment for future classes. Also we plan to modify this rubric using the same dimensions for other courses for which the dimensions adequately describe cognition.

References

1. *Rising Above the Gathering Storm : Energizing and Employing America for a Brighter Economic Future*. 2007, National Academy of Sciences: Washington, D.C; ASEE, *Engineering Education for the Global Economy: Research, Innovation, and Practice*, L.H. Jamieson and J.R. Lohmann, Editors. 2008: Washington DC; Besterfield-Sacre, M., L.J. Shuman, H. Wolfe, C.J. Atman, J. McGourty, R.L. Miller, B.M. Olds, and G.M. Rogers, *Defining the outcomes: A framework for EC-2000* IEEE Transactions on Education, 2000. **43**(2): p. 100-110; Seltzer, K. and T. Bentley, *The creative age: Knowledge and skills for the new economy*. 1999: Demos. 91.
2. Shuman, L.J., M. Besterfield-Sacre, and J. McGourty, *The ABET "Professional Skills" – Can They Be Taught? Can They Be Assessed?* Journal of Engineering Education, 2005. **94**(1).
3. Smith, K.A., S.D. Sheppard, D.W. Johnson, and R.T. Johnson, *Pedagogies of engagement: Classroom-based practices.* Journal of Engineering Education, 2005. **94**(1): p. 87-101; Golter, P.B., B.J. Van Wie, P.V. Scuderi, T.W. Henderson, R.M. Dueben, G.R. Brown, and W.J. Thomson, *Combining mordern learning pedagogies in fluid mechanics and heat transfer.* Chemical Engineering Education 2005. **39**: p. 280-287 ; Felder, R.M. and R. Brent, *Designing and Teaching Courses to Satisfy the ABET Engineering Criteria.* 2003: p. 7-25.
4. Streveler, R.A., T.A. Litzinger, R.L. Miller, and P.S. Steif, *Learning conceptual knowledge in the engineering sciences: Overview and future research directions.* Journal of Engineering Education, 2008: p. 279-294.
5. Streveler, R.A., B.M. Olds, R.L. Miller, and M.A. Nelson. *Using a Delphi Study to Identify the Most Difficult Concepts for Students to Master in Thermal and Transport Science*. in *American Society for Engineering Education Annual Conference and Exposition*. 2003: American Society for Engineering Education; Olds, B.M., R.A. Streveler, R.L. Miller, and M.A. Nelson. *Preliminary results from the development of a concept inventory in thermal and transport science*. in *American Society for Engineering Education Annual Conference & Exposition*. 2004. Salt Lake City, UT
6. Golter, P., B. Vanwie, G. Brown, D. Thiessen, and B. Abdul. *Shifting gears: Moving away from the controlled experimental model while improving rigor in engineering education research*. in *2010 American Society for Engineering Education Annual Conference and Exposition*. 2010. Louisville, Kentucky.
7. Golter, P., B. Vanwie, G. Brown, D. Thiessen, N. Yurt, and B. Abdul. *Aligning assessment tools with course subject and goals*. in *American Society for Engineering Education 2009 Annual Conference & Exposition* 2009. Austin Texas.
8. Hestenes, D. and I.A. Halloun, *Interpreting the force concept inventory.* The Physics Teacher, 1995. **33**(8): p. 502-506.
9. Zhao, Y., *How to design and interpret a multiple-choice-question test: a probabilistic approach.* International Journal of Engineering Education, 2006. **22**(6): p. 1281-1286.
10. Montfort, D., S. Brown, and D. Pollock, *An Investigation of Students' Conceptual Understanding in Related Sophomore to Graduate-Level Engineering and Mechanics Courses.* 2009: p. 111-129.
11. Marra, R.M., B. Palmer, and T.A. Litzinger, *The Effects of a First-Year Engineering Design Course on Student Intellectual Development as Measured by the Perry Scheme.* Journal of Engineering Education, 2000. **89**(1): p. 39-45.
12. Perry, W.G., *Forms of Ethical and Intellectual Development in the College Years.* 1999, San Francisco: Jossey-Bass
13. Stiggins, R.J., *Student-centered Classroom Assessment* Vol. 2. 1997, Gale: Prentice Hall.
14. Laeser, M., B.M. Moskal, R. Knecht, and D. Lasich, *Engineering Design: Examining the Impact of Gender and the Team's Gender Composition.* Journal of Engineering Education, 2003: p. 49-56; Jiusto, S. and D. Dibiasio, *Experiential Learning Environments: Do They Prepare Our Students to be Self-Directed, Life-Long Learners?* Journal of Engineering Education, 2006: p. 195-204; Charyton, C., R.J. Jagacinski, J.A.

Merrill, W. Clifton, and S. Dedios, *Assessing Creativity Specific to Engineering with the Revised Creative Engineering Design Assessment.* 2011. **100**(4): p. 778-799.

15. Greenberg, J.E., N.T. Smith, and J.H. Newman, *Instructional Module in Fourier Spectral Analysis, Based on Principles of "How People Learn".* Journal of Engineering Education, 2003: p. 155-165; Jonassen, D.H. and Y.H. Cho, *Fostering Argumentation While Solving Engineering Ethics Problems.* Journal of Engineering Education, 2011. **100**(4): p. 680-702; Carpenter, S.L., H.S. Delugach, L.H. Etzkorn, P.A. Farrington, J.L. Fortune, D.R. Utley, and S.S. Virani, *A Knowledge Modeling Approach to Evaluating Student Essays in Engineering Courses.* Journal of Engineering Education, 2007: p. 227-239; Colby, A. and W.M. Sullivan, *Ethics Teaching in Undergraduate Engineering Education.* Journal of Engineering Education, 2008: p. 327-338.
16. Yadav, A., G.M. Shaver, and P. Meckl, *Lessons Learned: Implementing the Case Teaching Method in a Mechanical Engineering Course.* Journal of Engineering Education, 2010: p. 55-69.
17. RomanTaraban, C. Craig, and E.E. Anderson, *Using Paper-and-Pencil Solutions to Assess Problem Solving Skill.* Journal of Engineering Education, 2011. **100**(3); Pandy, M.G., A.J. Petrosino, B.A. Austin, and R.E. Barr, *Assessing Adaptive Expertise in Undergraduate Biomechanics.* Journal of Engineering Education, 2004: p. 211-222; Taraban, R., A. Definis, A.G. Brown, E.E. Anderson, and M.P. Sharma, *A Paradigm for Assessing Conceptual and Procedural Knowledge in Engineering Students.* Journal of Engineering Education, 2007: p. 335-345; Litzinger, T.A., P. Vanmeter, C.M. Firetto, L.J. Passmore, C. B.Masters, S.R. Turns, G.L. Gray, F. Costanzo, and S.E. Zappe, *A Cognitive Study of Problem Solving in Statics.* Journal of Engineering Education, 2010(337-353).
18. Pellegrino, J.W., N. Chudowsky, and R.E. Glaser, *Knowing What Students Know : The Science and Design of Educational Assessment* 2001, Washington DC: National Academy Press.
19. Allen, S. and J. Knight, *A Method for Collaboratively Developing and Validating a Rubric.* International Journal for the Scholarship of Teaching and Learning, 2009. **3**(2): p. 1-17.
20. Ehrmann, S.C. (2011) *Rubrics: Definition, Tools, Examples, References*.
21. Yadav, A., G.M. Shaver, and P. Meckl, *Lessons Learned: Implementing the Case Teaching Method in a Mechanical Engineering Course.* Journal of Engineering Education, 2010. **99**(1): p. 55-69.
22. Mills, J.E. and D.F. Treagust (2003) *Engineering education-Is problem-based or project-based learning the answer?* Australasian Journal of Engineering Education, 1-15; Vanwie, B.J., D.C. Davis, P.B. Golter, A. Ansery, and B. Abdul. *Team building in a project-based learning class*. in *2011 ASEE Conference and Exposition*. 2011. Vancouver, BC: ASEE.
23. Fink, L.D., *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses* Jossey-Bass Higher and Adult Education. 2003, San Francisco, CA: Jossey-Bass. 287.
24. Litzinger, T.A., L.R. Lattuca, R.G. Hadgraft, and W.C. Newstetter, *Engineering Education and the Development of Expertise.* Journal of Engineering Education, 2011. **100**(1): p. 123-150.
25. Schachterle, L., C. Demetry, and J.A. Orr, *Quality Assurance in Engineering Education in the United States*, in *Engineering Education Quality Assurance: A Global Perspective*, A.S. Patil and P.J. Gray, Editors. 2009, Springer Science + Business Media, LLC New York. p. 170-175.
26. Brumm, T.J., S.K. Mickelson, B.L. Steward, and A.L. Kaleita, *Competency-based Outcomes Assessment for Agricultural Engineering Programs.* International Journal of Engineering Education, 2006. **22**(6): p. 1163.
27. Rugarcia, A., R.M. Felder, D.R. Woods, and J.E. Stice, *The future of engineering education I. A vision for a new century.* Chemical Engineering Education, 2000. **34**(1): p. 16-25.
28. Turns, J., C.J. Atman, R.S. Adams, and T. Barker, *Research on Engineering Student Knowing: Trends and Opportunities.* Journal of Engineering Education, 2005: p. 27-40.
29. Reeves, T.C., J. Herrington, and R. Oliver (2002) *Authentic activities and online learning*. **9**, 562; Bergen, D., *Authentic Performance Assessments.* Childhood Education, 1993. **70**; Terwilliger, J., *Semantics, Psychometrics, and Assessment Reform: A Close Look at "Authentic" Assessments.* Educational Researcher, 1997. **26** (8): p. 24-27.
30. Wiggins, G. (1990) *The case for authentic assessment. , . Retrieved October 11*. Practical Assessment, Research & Evaluation **2**.
31. Prince, M.J. and R.M. Felder, *Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases.* Journal of Engineering Education, 2006. **95**(2): p. 123.
32. Svinicki, M.D., *A Guidebook on Conceptual Frameworks for Research in Engineering Education*. 2008, NSF

33. Krathwohl, D.R., *A revision of Bloom's taxonomy: An overview* Theory into Practice 2002. **41**(4): p. 212-218; Ferris, T.L.J., *Bloom's Taxonomy of Educational Objectives: A Psychomotor Skills Extension for Engineering and Science Education.* International Journal of Engineering Education, 2010. **26**(3): p. 699-707.
34. Marques, J.F. and C. McCall, *The Application of Interrater Reliability as a Solidification Instrument in a Phenomenological Study.* The Qualitative Report Volume 2005. **10**(3): p. 439-462.
35. Nesbit, J., K. Belfer, and J. Vargo, *A Convergent Participation Model for Evaluation of Learning Objects.* Canadian Journal of Learning and Technology, 2002. **28**(3).
36. McCabe, W.L., J.C. Smith, and P. Harriot, *Unit Operations of Chemical Engineering*. 7th ed. McGraw-Hill Higher Education. 2005, New York: McGraw-Hill. 1140.

Appendices

Appendix A

Table A1: Current version of the FAI rubric used in this work

Score Levels – ***Significant Learning Anchor = 3 or B-***

	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
	F	**D**	**D+**	**C-**	**C**	**C+**	**B-**	**B**	**B+**	**A-**	**A**
Rubric dimension											
Foundational knowledge Recall basics i.e.: words/definitions/equations/ principles/concepts/relationships/logic/ understanding Show understanding via explanations, rationale for choices	No evidence of correct definitions/equations. Explanations incoherent obviously have no bearing on/ knowledge of FMHT principles.		Evidence of sketchy, disjointed explanations of FMHT principles; Equations incomcomplete/inconsistent with system Patchwork of correct ideas, sometimes contradictory (e.g., opposite ideas on the same principle in different parts of problem).		Evidence that explanations of FMHT principles appear correct but lack enough substance to be considered complete; Equations essentially complete but lack some significant terms		Evidence that explanations of FMHT are coherent and reasonable but lacking in finishing/completing touches or contain some slight error (mathematical or 'minor' conceptual flaws).		Equations / definitions complete and correct; FMHT explanations principles complete & perfect; thorough understanding of foundational principles as evident from explanations or usage of models.		
Application: Equation aligned with system /problem;	No evidence of ability to apply FM/ HT principles and /or		Vague evidence of how to apply FM and/or HT model, applica-		Some evidence of adequate grasp of model but applied incompletely;		Evidence of good ability in applying relevant models/ princi-		Evidence of complete & correct application; Appropriate mod-		

Equation reduced to conform with system; Reasonable assumptions; Correct use of industrial standards & tables e.g. schedule no., nominal and real size. Critical thinking- students analyze & evaluate correct units & sign; Creative thinking – outside box, new ways to solve; Practical thinking – good decisions, identifying practically infeasible solutions e.g. heat gain by cold fluid cannot be greater than that lost by hot.	models correctly; Model picked randomly without rationalization & no bearing to the problem; Assumptions are totally unreasonable; Deficient in critical, creative practical thinking.	tion is incorrect in substance & sequence; Terms irrelevant to system of interest included/ relevant terms omitted; Inappropriate parameters used; assumptions partially unreasonable; Solution incomplete; Sparse critical, creative or practical thinking.	Inappropriate parameters; Reasonable but incomplete assumptions, Significant errors in solution; Some errors in critical, creative or practical thinking.	ples/parameters; Slight errors like sign or unit consistency errors; Good critical, creative or practical thinking.	el/principle/parameters chosen, reasonable assumptions, flawless substitution & solution; Excellent critical, creative or practical thinking (e.g., when the measured heat gain by cold fluid is greater than the heat loss by hot fluid, students point that out as an error).
Integration: Correct integration of multi-transport principles (e.g. FM&HT, mass transfer & HT) i.e. the consequences of FM principles on HT understood, e.g., Re dictating form of Nusselt correlation, the direction of fluid flow dictating choice of Nusselt correlation, wet bulb temperature dictating amount of moisture in air.	No evidence of integration of mass transfer, FM&HT principles in a multi-transport problem; Models picked randomly (e.g. use of model appropriate for laminar to a turbulent flow FM&HT problem).	Evidence of application of mass transfer, FM & HT models / principles but no visible integration; (e.g., don't show that the flow regime determines the form of heat transfer coefficient to use, no explanation of how the various transport phenomena influence each other); solution is unrealistic.	Evidence of vague grasp of integration of mass transfer, FM&HT principles, but starkly incomplete: mixes unmatched correlations (e.g., correlations for turbulent & transition); choosing a correct correlation without or before(sequencing issue) calculating Re Assumptions unreasonable even though solution appears reasonable; mixed units.	Evidence of good integration of mass transfer, FM&HT models/principles; Appropriate mod-models/principles/solution sequence but with minor errors like sign, minor unit or conversion errors.	Evidence of excellent integration of mass transfer, FM&HT; Correct assumptions/ Parameters/sequencing of integration steps/ units and dimensions.

Appendix B. Typical assessment artifacts

Packed bed

Problem 7.3, McCabe, Smith and Harriott, 7ed [36]
The pressure drop for airflow through a column filled with 1-in. ceramic Raschig rings is 0.01 in. water per foot when Go = 80 lb/ft2·h and 0.9 in./ft water when Go = 800 lb/ft2·h, all for a mass velocity of the liquid flowing counter-currently of 645 lb/ft2·h. Since the change in pressure drop with liquid rate is slightly in the range of liquid mass velocities between 645 and 1,980 lb/ft2·h, ignore the liquid holdup and estimate the void fraction if the rings have a wall thickness of ⅛ in. Use this void fraction and the Ergun equation to predict the pressure drop, and discuss the difference between predicted and experimental values.

Fluidized bed

1. Answer the following questions
a. The Ergun equation represents the pressure drop as the sum of two terms. What do each of these terms represent physically?
b. Define each variable and constants in the equation.
c. What do the constants represent and how were they obtained?
2. Turn on flow from the tank and flow water up through the bed of glass beads and fill in the table below (take no more than 10 minutes on this):

Flow [GPH]	ΔP [in.H_2O]	Bed height, L [in.]	Observations [what you notice happening at each flow rate].

3. Before you leave the DLM please draw the fluidized bed and the differential pressure taps. What do you think will happen if the upper pressure tap was placed above the mesh?
4. After you are done with the DLM answer the following questions:
a. What happens when a bed becomes fluidized? Explain in terms of a balance of forces and other physical principles [a free body diagram could be helpful here].
b. Record the flow rate where you observed "fluidization". Calculate the corresponding superficial velocity and compare to that predicted by the appropriate equation in McCabe et al.
c. On the same axis, make plots of ΔP and L as functions of flow rate. Comment on and explain your observations.
d. From the relevant equation in McCabe and your measurements, (i) calculate the highest safest operating velocity for your bed without entrainment. (ii) Discuss the implications of increasing the particle size 10-fold (use a calculation to buttress your argument).
Column and packing information:

Column height 16 cm
Column ID 1.25 in. (31.75 mm)
Column OD 1.5 in. (38.1 mm)
Packing soda lime glass spherical beads (density of soda lime glass ~2.5 g/cm3)
Packing (spheres) diameter 0.8 – 1.2 mm