



Best 2019 Zone II Paper : Research to Practice: Leveraging Concept Inventories in Statics Instruction

Prof. Ruth Wertz P.E., Valparaiso University
Theresa Green, Utah State University - Engineering Education

Research to Practice: Leveraging Concept Inventories in Statics Instruction

Theresa Green, Valparaiso University
Ruth E. H. Wertz, Valparaiso University

Abstract

There are many common challenges with classroom assessment, especially in first-year large enrollment courses, including managing high quality assessment within time constraints, and promoting effective study strategies. This paper presents two studies: 1) using the CATS instrument to validate multiple-choice format exams for classroom assessment, and 2) using the CATS instrument as a measure of metacognitive growth over time. The first study focused on validation of instructor generated multiple choice exams because they are easier to administer, grade, and return for timely feedback, especially for large enrollment classes. The limitation of multiple choice exams, however, is that it is very difficult to construct questions to measure higher order content knowledge beyond recalling facts. A correlational study was used to compare multiple choice exam scores with relevant portions of the CATS assessment (taken within a week of one another). The results indicated a strong relationship between student performance on the CATS assessment and instructor generated exams, which infers that both assessments were measuring similar content areas. The second study focused on a metacognition, more specifically, on students' ability to self-assess the extent of their own knowledge. In this study students were asked to rank their confidence for each CATS item on a 1 (not at all confident) to 4 (very confident) Likert-type scale. With the 4-point scale, there was no neutral option provided; students were forced to identify some degree of confident or not confident. A regression analysis was used to compare the relationship between performance and confidence for pre, post, and delayed-post assessments. Results suggested that the students' self-knowledge of their performance improved over time.

Keywords: classroom assessment; validation; concept inventory

Introduction

There are many common challenges with classroom assessment, especially in first-year large enrollment courses. Statics is typically offered in the second semester of the first year, or first semester of the second year for engineering students. In addition, due to its applicability to multiple engineering disciplines, it typically has relatively high enrollments. Ideal assessments in statics would have students demonstrate procedural and conceptual knowledge of rigid body systems in multiple-formats, and verbally explain underlying concepts of procedural steps. Logistical constraints, and the need for timely feedback, generally prohibit such extensive assessment. On the other hand, multiple-choice exams, which can be very efficient for timely feedback, are very difficult to construct in such a way that they provide meaningful assessment of higher order cognition. In addition to assessment challenges, teaching first-year students is also challenging because of underdeveloped study skills. It is not uncommon for students to come into exams feeling overconfident because they have employed ineffective review techniques.

This work presents a series of two studies to address issues commonly faced by instructors who teach introductory mechanics courses, particularly within the first three years of teaching. The first study presented in the paper uses the *Concept Assessment Tool for Statics* (CATS) assessment to validate instructor written multiple-choice exams for summative assessment applications in a statics course.

The second study presented in this paper uses the CATS assessment along with an associated confidence scale to provide students with feedback on how the relationship between their confidence and performance changes over time. This evidence from prior years was shared with current students, along with recommendations for effective study strategies.

Literature Review

Classroom Assessment in Statics,

Multiple-choice exams are widely recognized as a convenient method of assessing student understanding. They are particularly useful in large classrooms where administering constructed-response exams would be logistically infeasible. Multiple-choice exams are an ideal test format for large enrollment courses since less time and effort is required to grade and redistribute them to students, compared to constructed-response exams (Scott et al., 2006). National exams such as advanced placement (AP), the Scholastic Assessment Test (SAT), and the graduate record examination (GRE) use the multiple-choice format for this reason. However, the ability of multiple-choice exams to accurately and fully assess student knowledge and critical-thinking skills is under debate (Simkin & Kuechler, 2005; Stanger-Hall, 2012). Simkin and Kuechler (2005) argue that while some multiple-choice questions do test only superficial knowledge, questions can be designed such that they target different levels of student learning and understanding. While multiple-choice exams are a logical way for a large number of exams to be implemented and graded, it is difficult for instructors to design multiple-choice questions that target concepts requiring a deeper level of understanding, as opposed to strictly remembering facts. This challenge suggests the need for validating multiple-choice exams that claim to target these complex concepts. One way to potentially validate an instructor-created multiple-choice

exam is through an external measurement from an exam or assessment that has already been validated. One such exam in the engineering domain is the Concept Assessment Tool for Statics (CATS) (Steif & Dantzler, 2005).

The CATS is a multiple-choice concept inventory designed to measure conceptual knowledge and target student misconceptions about topics in engineering statics. This instrument can be used by instructors to understand where students are struggling conceptually, allowing them to adjust their teaching methods and lessons appropriately (Steif & Dantzler, 2005). Steif and Dantzler (2005) verify that the CATS assessment does indeed measure students' conceptual knowledge in statics through three levels of content validity, construct validity, and criterion-related validity. Their study showed significant, positive correlations between inventory score and course grade, indicating that "the total score of the inventory is a valid measure of statics conceptual knowledge." Since CATS scores provide a valid and reliable measure of student understanding and conceptual knowledge, we use the assessment in our study.

The purpose of the first study was to observe whether there is a correlation between scores on the class exams and scores on the CATS assessment. Our results may indicate that exam validity can be inferred through the use of an external measurement, such as the CATS, and also reveal that the class exams test the same level of conceptual knowledge as the CATS.

The Role of Metacognition in Conceptual Change

Conceptual knowledge is one's collection of understanding and intuition that provides a foundation of competence and familiarity within a subject or idea. Engineering practice relies on the successful utilization of conceptual knowledge with regard to system design, problem-solving processes, and developing professional competence in the field (Streveler & Litzinger, 2008). Current research investigates both methods of targeting change in conceptual knowledge (Chi, 2008) and methods of assessing conceptual knowledge (Streveler & Litzinger, 2008) in the engineering sciences. Measuring conceptual knowledge and how it changes allows instructors to have a better understanding of the areas in which students struggle. One method of measuring conceptual knowledge is through the use of concept inventories, because they are a fast method of assessment and have been developed for numerous areas in science and engineering (Streveler & Litzinger, 2008). For engineering specifically, the Statics Concept Inventory (Steif & Dantzler, 2005) was developed to identify and target student misconceptions about common topics in statics. A concept inventory for statics is useful since the course is pertinent in the engineering curriculum; many courses build upon both the content and problem-solving strategies developed in a statics course (Steif & Dantzler, 2005).

Students may have different degrees of conceptual knowledge when entering the classroom, and instructors must design their curriculum to account for varying misconceptions accordingly (Streveler & Litzinger, 2008). There are different ways to design instruction such that these misconceptions are corrected, depending on the way the students' knowledge is misconceived (Chi, 2008). Recognizing the differing conceptual frameworks that students bring to a class allows the instructor to tailor their methods to help rebuild students' misconceptions into a correct conceptual framework, thus prompting conceptual change (Streveler & Litzinger, 2008).

Conceptual change is recognized as a type of learning where prior misconceptions about a concept are changed into correct knowledge through learning, and can be achieved by correcting

“false beliefs” about a topic through “belief revision” (Chi, 2008). Belief revision occurs when false beliefs are directly refuted with the correct information (Chi, 2008). One approach to encourage conceptual change in engineering statics in particular is through metacognitive prompts: questions designed to focus student attention on planning and evaluating different ways of representing a problem before beginning a solution (Steif, Lobue, Kara, & Fay, 2010). Steif et al. (2010) implemented metacognitive prompts in the form of “body-centered talk” in an introductory statics course, where students were prompted to focus on key concepts and develop a conceptual understanding of the problem before attempting to solve it. The “body-centered talk” focused on encouraging students to think specifically about bodies in the problem, interactions between bodies, and drawing correct free-body diagrams. Their results show that students’ problem-solving performance in statics may improve if students are prompted to think about the underlying concepts in the problem before immediately beginning a solution, because they must recognize and apply the appropriate conceptual knowledge. Additionally, students that have more developed conceptual knowledge may have enhanced problem-solving abilities because they have the ability to understand the problem on a deeper level, and they can more easily identify and apply unique ways to solve problems (Streveler & Litzinger, 2008).

Pintrich (2002) advocates that accurate self-knowledge is essential for learning. If students have an inflated sense of their own understanding, they are less likely to expend effort to acquire better skills or construct new knowledge. The CATS instrument by itself is an assessment only of conceptual understanding, but paired with a self-reported confidence level, we can begin to understand how well students are able to identify what they do and do not know.

The purpose of the second study was to further investigate the use of the CATS instrument, paired with self-reported confidence, to understand how students’ self-knowledge developed over time.

Methods

The participants in this study consisted of a group of 105 engineering students distributed over three sections, each taught by the same professor, in a statics course at a small Midwestern university. There were 21% female students and 81% male students enrolled in the course, which is representative of the overall enrollment in the College of Engineering.

The CATS (Steif & Dantzler, 2005) instrument was used to measure conceptual understanding in a pre, post, and delayed post research design. All 27 items of the CATS instrument were imported to an online survey environment. The survey was designed to show students one item at a time, which was paired with a confidence scale for each item. The confidence scale was a Likert-type scale where 1 = very uncertain (i.e. equivalent to a random guess), 2 = uncertain (i.e., student could identify at least one option as incorrect), 3 = certain, and 4 = very certain. Note that a neutral confidence option was not offered; students had to identify as either certain or uncertain to some degree. Completing the single item and the associated confidence interval were both required before the student could proceed to the next item; backtracking was also prohibited. The 27 items are grouped by concept as described in Table 1.

The pre-assessment was conducted online on the first day of class as a set of all 27 items. The statics course was taught as a series of five modules, and the post-assessments were thus

disseminated online as subsets of three to nine items throughout the semester based on the topics covered in each module. The delayed post-assessment was administered in a 50-minute class period at the end of the term. Table 2 summarizes which items of the CATS assessment were disseminated during the pre, post, and delayed-post assessment periods.

Table 1: CATS Concepts by Item

Concepts	Description	Items
Forces on collection of bodies	Identifying forces acting on a subset of a system of bodies	Set 1: 1-3
Newton's 3 rd law	Forces between two contacting bodies must be equal, opposite, and collinear.	Set 2: 4-6
Static equivalence	Static equivalence between forces, couples, and combinations	Set 3: 7-9
Roller connections	Direction of force between roller and contacting surface	Set 4: 10-12
Pin and slot connections	Direction of force between pin and slot member	Set 5: 13-15
Negligible friction	Direction of force between frictionless bodies at point of contact	Set 6: 16-18
Representation	Representing unknown loads at various connections	Set 7: 19-21
Friction	Combining equilibrium and Coulomb's Law for friction forces	Set 8: 22-24
Equilibrium	Considerations of both force and moment balance in equilibrium	Set 9: 25-27

Steif, P. (2010). Concept Assessment Tool for Statics: Concepts/Examples. Retrieved from <https://engineering-education.com/CATS-concepts.php>.

Table 2.0 Data Collection Summary

Module Topic	Hibbeler Chapters ¹	CATS Items		
		Pre	Post	Delayed Post
M1: Particle Equilibrium	2, 3	1-27	1-3, 4-6	1-27
M2: Rigid Body Equilibrium	4.1-4.5, 5	1-27	10-12, 13-15, 25-27	1-27
M3: Equivalent Systems, Centroids, and Moment of Inertia	4.6-4.9; 9.1-9.2; 10.1, 10.2, 10.4	1-27	7-9	1-27
M4: Analysis of trusses, frames, and machines	6	1-27	16-18, 19-21	1-27
M5: Internal forces and static friction	7.1, 7.2; 8.1, 8.2	1-27	22-24 ²	1-21, 25-27

Notes: 1. Hibbeler, R. C. (2016). Engineering Mechanics: Statics, 14th Edition. Pearson – Prentice Hall. 2. There is no delayed-post measurement for CATS Set 8, which includes items 22, 23, and 24.

Results

Study #1: Validation of Instructor-Generated Exams

A correlational study of the CATS post items with each corresponding exam, as well as the delayed post items with the final exam was performed using SPSS. The results (summarized in Table 2) indicated a statistically significant correlation between student performance on the CATS assessment, and student performance on each exam, except for Module 5. This likely due to the limited time students have to learn the content in Module 5, which is only covered in the last week and a half (4 class periods) before the finals exam.

Table 3: Pearson's correlations between CATS and Exam performance by module

	CATS1	CATS2	CATS3	CATS4	CATS5	CATS DP
	N=96	N=98	N=88	N=97	N=98	N=98
Exam M1	.263**	-	-	-	-	-
Exam M2	-	.382**	-	-	-	-
Exam M3	-	-	.290**	-	-	-
Exam M4	-	-	-	.326**	-	-
Exam M5	-	-	-	-	.150	-
Exam Final	-	-	-	-	-	.516***

Notes: * significant at $p < 0.05$; ** significant at $p < 0.005$; *** DP = delayed post; M5 is a subset of the final exam

The number of participants (N) taking the module exams and the final exam differed for each exam. The number of students for each assessment may vary based on absentees or withdraws. For each paired t-test, missing values were eliminated pairwise, so N represents the number of cases with complete exam scores and post assessments.

Study #2: Relationships between Student Confidence and Student Performance over Time

Simple linear regression was used to evaluate to what extent student performance was predicted based on their self-reported confidence for each item. Regression analyses were completed for each set of items representing a single topic, with nine analyses total. The analyses were then repeated for the post-test and delayed post-test. The standardized regression analysis results are presented in Table 4 for the pre-test data, Table 5 for the post-test data, and Table 6 for the delayed post-test data.

Table 4: Results of Standardized Regression Analyses by Topic (Pre-Test)

Predictors	Beta	t	Sig (p)	F	df	R ²
Ave Confidence Set 1	-.14	-1.41	.163	1.98	1, 101	.02
Ave Confidence Set 2	.22	2.23	.028	4.99	1, 101	.05
Ave Confidence Set 3	.09	.14	.888	.02	1, 101	.00
Ave Confidence Set 4	.35	3.80	.000	14.43	1, 101	.13
Ave Confidence Set 5	.19	1.98	.051	3.91	1, 101	.04
Ave Confidence Set 6	-.21	-2.20	.031	4.85	1, 101	.05
Ave Confidence Set 7	.27	2.76	.007	7.64	1, 101	.07
Ave Confidence Set 8	-.00	0.03	.976	.00	1, 101	.00
Ave Confidence Set 9	.20	2.00	.048	4.01	1, 101	.04

Note. The dependent variable for each analysis is the CATS score for each respective set.

Table 5: Results of Standardized Regression Analyses by Topic (Post-Test)

Predictors	Beta	t	Sig (p)	F	df	R ²
Ave Confidence Set 1	.20	1.95	.054	3.80	1, 95	.04
Ave Confidence Set 2	-.03	-.30	.768	.09	1, 96	.00
Ave Confidence Set 3	.39	3.89	.000	15.09	1, 85	.15
Ave Confidence Set 4	.43	4.60	.000	21.14	1, 95	.18
Ave Confidence Set 5	.37	3.80	.000	14.41	1, 93	.13
Ave Confidence Set 6	.19	1.89	.063	4.85	1, 93	.04
Ave Confidence Set 7	.47	5.12	.000	26.23	1, 95	.22
Ave Confidence Set 8	.14	1.38	.171	1.93	1, 95	.02
Ave Confidence Set 9	.30	2.99	.004	8.93	1, 93	.09

Note. The dependent variable for each analysis is the CATS score for each respective set.

Table 6: Results of Standardized Regression Analyses by Topic (Delayed Post-Test)

Predictors	Beta	t	Sig (p)	F	df	R ²
Ave Confidence Set 1	.54	6.27	.000	39.35	1, 95	.29
Ave Confidence Set 2	.25	2.51	.014	6.32	1, 95	.06
Ave Confidence Set 3	.36	3.78	.000	14.32	1, 95	.13
Ave Confidence Set 4	.50	5.61	.000	31.51	1, 95	.25
Ave Confidence Set 5	.23	2.35	.021	5.51	1, 95	.06
Ave Confidence Set 6	.13	1.27	.208	1.61	1, 95	.02
Ave Confidence Set 7	.42	4.53	.000	20.48	1, 95	.18
Ave Confidence Set 8	--	--	--	--	--	--
Ave Confidence Set 9	.33	3.46	.001	11.94	1,95	.11

Note. The dependent variable for each analysis is the CATS score for each respective set.

Discussion

Study #1: Validation of Instructor-Generated Exams

The significant correlations, varying from $r = 0.263$ to $r = 0.516$ ($p \leq 0.05$), between module exam scores and CATS scores demonstrate that the same level of understanding was being tested across both assessments. Both examinations were designed to measure students' conceptual understanding of the material presented in class. Since the CATS assessment has been previously validated as a reliable measure of conceptual knowledge in statics (Steif & Dantzler, 2005), the significant, positive correlation between the module exam scores and the CATS scores indicates that the module exams are also a valid measure of similar or related concepts. Overall, based on the data presented in Table 3, we can infer that the instructor-authored exams represented a reasonable summative measure of statics concepts.

Our results corroborate the proposed research goal. We sought to assess the validity of instructor-created multiple-choice exams in an engineering statics class using the CATS assessment as an external measure. The results of our study provide an example of effectively using external measurement tools as a way to validate classroom assessments. Validating classroom exams provides affirmation to the students that their exam grades are meaningful, and to the instructor that the exams are testing over the concepts that they claim it tests. It is also advocates the creation of multiple-choice exams in order to test higher level thinking beyond recalling facts, even though it is generally difficult to do so.

All analyses have their limitations. In the context of this investigation, the instructor generated exams accounted for a total of 60% of the students' final grades, and required students to show all computational work to receive credit (i.e., a correct answer with no work would be graded as incorrect). The CATS assessments were graded based on completion only, and accounted for approximated 5% of students' final grade. This imbalance may have led students to put more thought and effort into one assessment over the other and diminish the overall strength of the

relationship between the two. This, and the fact that the instructor generated exams included both conceptual and procedural problems, while the CATS assessment only included conceptual problems, would both tend to decrease (not increase) correlational values. Therefore, we still feel the results of this study, despite its limitations, are of value and interest to the community.

Study #2: Relationships between Student Confidence and Student Performance over Time

The regression results show that the extent to which students' self-reported confidence predicted performance on the CATS assessment varied both by time and by topic. Looking broadly, students' self-knowledge seems to improve over time. For the pre-test assessment, students' self-reported confidence significantly predicted performance in five of the nine CATS topics (sets 2, 4, 6, 7, and 9). Set 6 of this group (negligible friction) had a negative relationship where students with higher self-reported confidence were less likely to answer items correctly. By the post-test assessment students' self-reported confidence still significantly predicted performance in five of the nine CATS topics (sets 2, 3, 4, 7, and 9), but all relationships were positive and the percent of overall variance in performance predicted by self-reported confidence increased from the pre-test ($R^2 = 0.05$ for Set 2, $R^2 = 0.13$ for Set 4, $R^2 = 0.05$ for Set 6, $R^2 = 0.07$ for Set 7, and $R^2 = 0.04$ for Set 9) to the post-test ($R^2 = 0.15$ for Set 3, $R^2 = 0.18$ for Set 4, $R^2 = 0.13$ for Set 5, $R^2 = 0.22$ for Set 7, and $R^2 = 0.09$ for Set 9). By the delayed post-test assessment students' self-reported confidence significantly predicted performance for all but set 6 (negligible friction), which is the same set that started with a negative relationship between confidence and performance.

The course work included activities specifically designed to target metacognitive knowledge. These activities included regularly classifying key features of the system (coplanar vs. 3D, concurrent or non-concurrent, force or force couple combination, etc.) and classifying problem type (system definition, equivalent system, equilibrium, force analysis, or equilibrium with friction. Students were also prompted to use the open source OLI Engineering Statics course for self-testing at regular intervals throughout the term. Students were not required to officially enroll in the online course so usage statistics were not available, but students were required to provide screenshots of completed tutorials for critical topics, such as generating free body diagrams (Sets 1, 4, 5, and 7), effects of force (Set 3 and Set 9). Topics around friction were not addressed until the very end of the term, which may be why self-knowledge for these topics (Set 6 and Set 8) were not as well developed as others.

References

- Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. *Advances in Child Development and Behavior*, 10(C), 103–152. [http://doi.org/10.1016/S0065-2407\(08\)60009-9](http://doi.org/10.1016/S0065-2407(08)60009-9)
- Chi, M. T. H. (2008). Three types of Conceptual Change: Belief Revision, Mental Model Transformation, and Categorical Shift. *International Handbook of Research on Conceptual Change*, 61–82.
- Jorion, N., Gane, B., DiBello, L., & Pellegrino, J. (2015). Developing and Validating a Concept Inventory. In *2015 ASEE Annual Conference and Exposition Proceedings* (pp. 26.497.1–26.497.12). Seattle: ASEE Conferences. <http://doi.org/10.18260/p.23836>

- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R. (2002). The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing. *Theory Into Practice*, 41(4), 219–225.
http://doi.org/10.1207/s15430421tip4104_3
- Scott, M., Stelzer, T., & Gladding, G. (2006). Evaluating multiple-choice exams in large introductory physics courses. *Physical Review Special Topics - Physics Education Research*, 2(2), 020102–1 – 020102–14. doi: [10.1103/PhysRevSTPER.2.020102](https://doi.org/10.1103/PhysRevSTPER.2.020102)
- Simkin, M. G., & Kuechler, W. L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73–98. doi: [10.1111/j.1540-4609.2005.00053.x](https://doi.org/10.1111/j.1540-4609.2005.00053.x)
- Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: development and psychometric analysis. *Journal of Engineering Education*, 94(4), 363–371. doi: [10.1002/j.2168-9830.2005.tb00864.x](https://doi.org/10.1002/j.2168-9830.2005.tb00864.x)
- Steif, P. S., & Hansen, M. A. (2007). New Practices for Administering and Analyzing the Results of Concept Inventories. *Journal of Engineering Education*, 96(3), 205–212. doi: [10.1002/j.2168-9830.2007.tb00930.x](https://doi.org/10.1002/j.2168-9830.2007.tb00930.x)
- Steif, P. S., Lobue, J. M., Kara, L. B., & Fay, A. L. (2010). Improving Problem Solving Performance by Inducing Talk about Salient Problem Features. *Journal of Engineering Education*, 99(2), 135–142. <http://doi.org/10.1002/j.2168-9830.2010.tb01050.x>
- Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE Life Sciences Education*, 11(3), 294–306. doi: [10.1187/cbe.11-11-0100](https://doi.org/10.1187/cbe.11-11-0100)
- Streveler, R. A., Litzinger, T. A., Miller, R. L., & Steif, P. S. (2008). Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions. *Journal of Engineering Education*, 97(3), 279–294. <http://doi.org/10.1002/j.2168-9830.2008.tb00979.x>
- Weinstein, C. E., Schulte, A., & Palmer, D. (1987). *Learning and study strategies inventory*. Clearwater, FL: H & H Publishing.