The Development Process for a New Materials Science Conceptual Evaluation

Rebecca Rosenblatt and Andrew F. Heckler
Illinois State University rjrosen@ilstu.edu and The Ohio State University heckler@mps.ohio-state.edu

Abstract - **In this work, we introduce the Materials Science Conceptual Evaluation (MSCE). This is a 31 item multiple choice test designed to complement the Materials Concept Inventory (MCI) developed in 2002 by Krause et al. In developing this assessment, we collected data from over 700 students, including over 100 interviews, focused on areas where students struggle or have misunderstandings. There are several main goals of this article. The first is to inform the engineering education community of the existence of this evaluation and its psychometric properties, which support its validity and reliability as an assessment. The second goal is to describe our design and validation process and what we discovered to be the most, and least, important general steps for future engineering education assessment design and validation. Lastly, we will communicate some of the key findings about student difficulties with basic concepts and the required reasoning and graphical skills for an introductory materials science course.**

Index Terms – Conceptual difficulties, Assessment Design, Materials Science

I. INTRODUCTION

In this work, we introduce the Materials Science Conceptual Evaluation (MSCE). This is a 31 item multiple choice test designed to complement the Materials Concept Inventory (MCI) developed in 2002 by Krause et al. [1]. The test is very broad in scope and designed to cover the majority of topics for an introductory materials science and engineering course. Some example topics that the test assesses are: atomic bonding, atomic weight, crystal structure, diffusion, material properties, stress-strain diagrams, phase diagrams, creep, fatigue, and TTT plots. Because of this breadth, the term 'conceptual evaluation' is used. The test is not designed to be used as a concept inventory, which would have a greater variety of questions on a single topic. The test is designed to assess student understanding of a range of topics commonly covered in an introductory undergraduate materials science course.

There are three main goals of this article. The first is to inform the engineering education community of this evaluation, its psychometric properties, and its measures of validity such as correlations with final exam score, grade in the course, and years in the program.

The second goal is to provide details about the item design and validation process. This will allow instructors to confidently use this instrument or items from this instrument to assess their students and/or their instructional reforms. It will also provide a guide for other instructors if they wish to develop additional items and/or assessments.

The third goal is to communicate some of the key findings about student difficulties with basic concepts and the required reasoning and graphical skills for an introductory materials science course. In developing this assessment, data was collected from over 700 students, including over 100 interviews, focusing on areas where students have misunderstandings. Many of these areas of difficulty are already known in the existing body of research in materials science education, but some are new findings from this project. Through a presentation of student responses to items and student reasoning, we will report these findings and the insights they give into how to best teach introductory materials science. All of the data presented here was collected in an introductory materials science course for second and third year engineering majors.

II. OUR ITEM DESIGN PROCESS

A. *A Generally Accepted Design Process*

A well-established process for test and instructional material development was proposed in Beichner's 1994 paper [2, 3, 4]. These steps are: 1. Identify a need for a new test; 2. Determine goals for the test; 3. Create the items; 4. Conduct preliminary testing for validity and reliability with revisions to the items as needed; 5. Perform validation and reliability checks on the final items; 6. Distribute and use the test.

While this process is well known, and commonly used, there are many levels of rigor and precision with which steps 3-5 are completed, and there are many different ways that validity and reliability are established and reported [2, 6 - 10]. In this article, we present our process, which we believe is both rigorous and precise.

B. *Steps 1 & 2: The Need for a New Test*

The Materials Science Conceptual Evaluation emerged from several years of study on student difficulties with introductory materials science concepts [11-15] and was used to assess student learning from an instructional reform creating group work tutorials for use in the introductory materials science course [11]. The MCI, the only other

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

materials science assessment available at the time, did not have the detail in some areas to effectively assess the planned instructional change and did not address several of the areas of difficulty we observed.

C. Step 3: Initial Item Construction

There were several ways that new items were created for the MSCE and several ways that items were redesigned once they were created. The main contributors to the identification of new areas or topics of student difficulty were: existing literature on student difficulties, exploratory interviews with students, old exam questions, and student tutorial responses both written and verbal.

Once a topic or area was identified, subsequent student interviews and open ended test questions assessed the specifics of that difficulty and provided an estimation of the number of students who exhibited that difficulty. A multiple-choice question was then created using common mistakes exhibited by the students in their short-answer responses as distractor options.

While distractors are not part of classical test design theory, they are often used in science assessments because of their usefulness as an informative teaching tool. Distractors are options that are representative of common incorrect student responses to the item, and they are usually created as described above from student responses to the item when given as an open ended question, e.g. MCI, Force Concept Inventory, Baseline Electricity and Magnetism Assessment, and Force and Motion Concept Evaluation [1, 5, 16, 17].

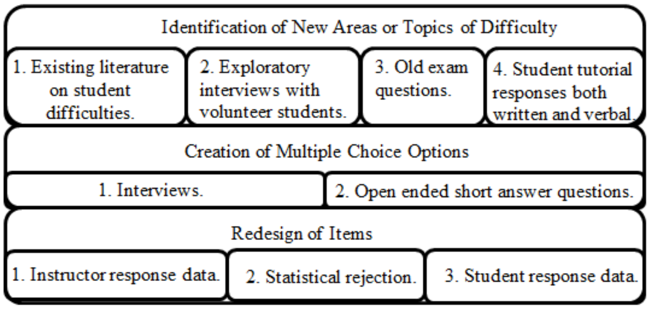


Figure1: The main contributors to each stage of the item design and re-design process

III. STEP 4: THE ITEM REDESIGN, VALIDATION, AND REJECTION PROCESS

Once an item was created, it went through a trail run with a small number of students. During this time many items were partially validated by requiring students to both select an answer and then give a short written response explaining their thought process. This was used to assess each item's ability to group students into their actual idea state. Subsequent rejection or redesign of these items was based on three main factors: instructor response data, student response data, and statistical rejection. All but a few of the MSCE test items were created in this fashion.

A. Theory of Validation and Reliability

Test reliability and validity are necessary for stable and applicable inferences to be drawn from a test. Validity and reliability, while commonly discussed separately, are very closely related to each other. In practice, a test which is not reliable is not valid and vice versa.

Test reliability is a measure of the error associated with the reported score on the test. There are three main types of reliability: internal consistency; stability; and equivalency [3]. Internal consistency is how related all the questions on the test are to each other. It assumes the test assesses only a single trait x and thus correctness on all questions should be correlated; this can be measured by a Kuder-Richardson -20 (KR20). Stability addresses the likelihood that the same score would be received by a student if he or she could retake the test; consistency of student scores between semesters indicates this. Equivalency is how much agreement there is between the test's measure of trait x and a separate set of *equivalent* items' measure of trait x; correlations with other measures of student knowledge such as exam scores and tutorial responses, indicate this.

Test validity is usually defined as the extent to which a test measures what it is supposed to measure. As such, there is no single statistical measure of validity of a test. Thus, validity is established through a variety of measures. Contemporary thinking distinguishes three different types of validity. Content validity measures a tests ability to assess the materials it is supposed to be testing. This type of validity is most often completed by consultation with experts in the assessed topic. Construct validity measures a test's ability to assess a theoretical trait in the test taker. In our research, this is student understanding of a concept. Criterion validity measures a test's ability to assess transferable skills; in other words, to what extent does performance on the test relate to performance in other domains [3, 18, 19].

B. Instructor Response Data

The first way that items were reworded or removed was instructor response data. Three recent course instructors, who were faculty in the materials science department, analyzed the MSCE.0 items for: correctness, the importance of inclusion of each question on the MSCE test, and concern over students' response percentages. (These last two, importance and concern, were Likert rankings 0 to 10 and were used as an indicator of instructors' views of item importance and value.) Any questions where these three things were not highly rated by a majority of the instructors were reevaluated and either removed or reworded. Additionally, items were reevaluated each quarter by a group of four or five teaching assistants for the course. Changes to question wordings were made based on their suggestions. This process established the content validity.

Table I: A summary of MSCE test Versions and data collected on each

Version	# of items	N	Tutorials	Attendance	Recitation graded	Avg. Score	Std.	KR20	Correl. w/ Final Grade
MSCE.0	43	47	Not used	30% (est.)	Ungraded	44%	12%	0.562	0.290
MSCE.1	32	102	Used	36%	Ungraded	60%	16%	0.745	0.387
MSCE.2	32	221	Used	62%	Ungraded	56%	18%	0.803	0.623
	32	155	Not used	30% (est.)	Ungraded	44%	15%	0.735	Unavailable
MSCE.3	31	199	Used	93%	Graded	68%	17%	0.791	0.443

C. *Student Response Data*

Student interview and reasoning data was the main way that questions' construct validity was established and the second way for items to be rejected or reworded. Think-aloud interviews as students completed the test questions and question sets of 'choose-an-answer and then explain-your-choice' were given. In both cases these validation tests were done before, during, and after item creation. With the before and during used to edit the items and the after used as a validation measure for the final version.

D. *Statistical Rejection*

Statistical rejection was the third way items were chosen to be removed. Items with low or negative point biserial coefficients were eliminated. Items with difficulties that were either too high or too low were also removed.

Examples of questions which were statistically rejected are shown in Figure 2 and Figure 3. The first of these was simply too easy. While given an average score of 9 (out of 10) on inclusion in a concept evaluation, the item was giving no real indication of student learning. The second question the point biserial of -0.280 shows that students who did well on the test did not do well on this question.

Figure 2: A too easy question (N= 47, STE = 4%).

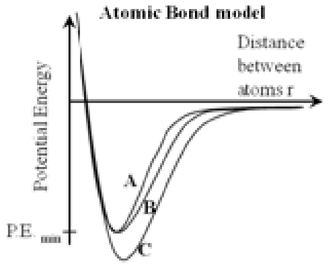
Which of the following will increase the diffusion rate of atoms in a metal?

90%.	Increase the temperature of the metal.
5%	Increase the activation energy of the metal.
2%	Increase the size of the atoms.
3%	None of the above will necessarily increase the diffusion rate.

Figure 3: Example of a statistically rejected item. (point biserial of -0.28, N = 47, STE = 5%).

Shown below are the Potential Energy curves for three different materials. Circle the answer with the correct rankings of the three materials' stiffness (i.e. their Young's moduli, E).

18%	A > B > C
30%	C < B = A
41%	C > B > A



E. *Summary of Version changes*

MSCE.0: The 43 items, which had been created through an iterative process of design and validation through interviews and open ended testing, were given to a small population of students and to three faculty to evaluate for content and construct validity.

MSCE.0 based on faculty evaluations and statistical reliability and validity. It had 32 items.

MSCE.1: This version is a subset of the questions of version

MSCE.2: This version included changes to four questions' wording. Also, two questions were exchanged for two different questions which were more central to the final tutorial course material. It had 32 items.

MSCE.3: This is the newest, and final, version of the MSCE. Small edits were made for clarity to the diffusion, phase diagram, and creep questions. It had 31 items.

IV. STEP 5: ESTABLISH THE VALIDITY OF THE MSCE

The criterion validity of the test was completed through statistical comparisons of the response data with other measurements of conceptual understanding or student skill. In the following sub-sections we will present several comparisons including: positive correlations of MSCE and final exam scores, positive correlations of MSCE with tutorial use (previously shown to be an effective teaching tool [11, 12]), higher scores when students put more effort into the test, and higher scores for the more experienced TAs. In no case did a comparison suggest that the test was invalid. All expected positive correlations were observed.

Table I reports the number of items on each test version, the population tested with each version, and the different tests' average score, standard deviation, reliability (KR 20), and correlation with final exam grade. There are no statistical abnormalities in this data that would suggest invalid tests. The averages on the different tests were neither at floor nor ceiling for any of the different student groups. The reliability for all but the MSCE.0 was greater than 0.7, and given the previously discussed status of the test as a concept evaluation, it would not be expected to have an especially high inter-item reliability because the items are testing a range of topics. In addition, each quarter shows reasonable correlations with grade on the final exam, which is a different measure of student ability in materials science.

A. *Gains in Item and Test Score with Tutorial Use*

Tutorials are an instructional reform that has previously been proven as an effective instructional tool [11, 12]. Thus we would expect higher scores in classes where tutorials are used. While different versions of MSCE were given in different classes as the test was developed, an analysis of the available data shows that students do better on the MSCE with tutorial use in class.

For example, the MSCE.2 was given as a posttest in two course sections, one section that employed tutorials during non-graded recitations (that is, attendance was not graded) and one that did not employ tutorials. In the section where tutorials were not used, the course format was identical, but the recitations were used for mini-lectures and homework help. The average score on the MSCE.2 was 47% without tutorials and 56% with tutorials. So, students scored 9%, or 0.53 standard deviations, better on the MSCE.2 in the quarter the tutorials were used ($t(360) = 24.10$, $p < 0.001$).

In addition, Figure 4 shows increased student score with increased tutorial use. The MSCE.1 and the MSCE.2 were both given in quarters (12 week courses) with nongraded recitation attendance. This led to variable attendance and thus variable exposure to tutorials. Figure 4 shows there is a linearly increasing relationship (in both quarters) between recitation attendance – i.e. tutorial use – and MSCE score. For each recitation attended, there was on average a 0.12 standard deviation increase in score on the MSCE.2 and a 0.18 standard deviation increase in score on the MSCE.1

Another interesting result, which is shown in Figure 4, is an increase in students' scores when all students are "required" to attend recitation. MSCE.3 was given in a quarter where recitation attendance was graded which resulted in 93% attendance, i.e. most students attended at least eight of the recitations. The average score of 67% for this quarter places the data in a near perfect fit with previous finding of tutorial use and increasing scores on the MSCE test.

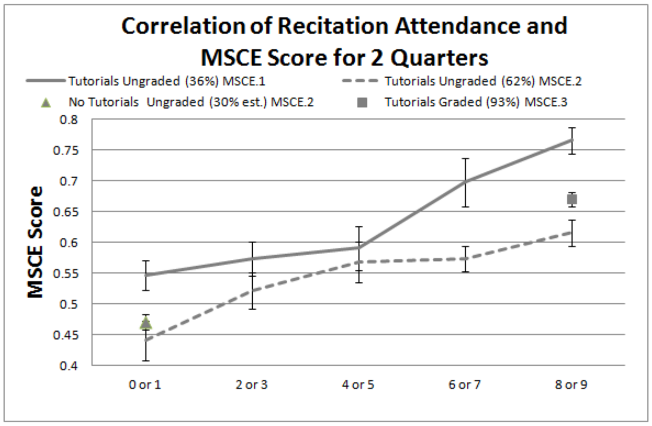


Figure 4: Gains in each MSCE version's score as a function of tutorial exposure. For each recitation attended, there was a 0.12-0.18 standard deviation increase in score.

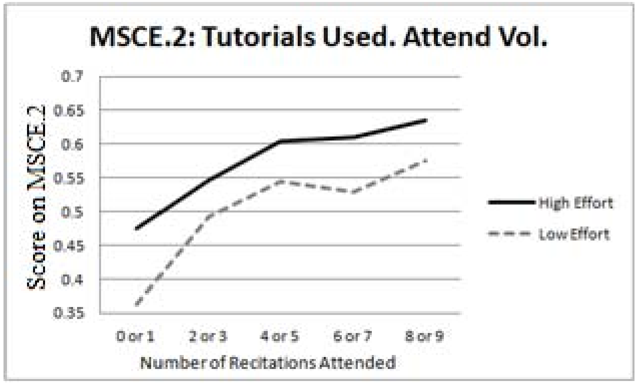


Figure 5: Tutorial use and effort on MSCE score

B. *MSCE Score Increase with Student Testing Effort*

MSCE tests were all administered to students in a special session in which they received participation credit. Students were told that their score on the MSCE test would not affect their grade in the course. In addition, students had no reason to study or review old material prior to the assessment. They were encouraged to try their hardest and treat the test as a practice final in which they could assess their preparedness for the upcoming final. In principle, this resulted in most of the students completing the test and giving a good faith effort. However, to further test the effect of effort, a posttest exit survey question asked students to rank from 1 to 5 their effort on the assessment. 1 was, `I guessed randomly on most of the questions,' and 5 was, `I tried my hardest and took an educated guess when unsure of the answer'.

This revealed two things. First, the average response value was high, 3.7, so students mostly reported good effort. Second, students who reported that they tried harder scored higher on the assessment. When the data is binned as high effort - 5 or 4 - and low effort - 3 or less - , students who reported high effort scored an average of 7.5%, or 0.44 standard deviations, better on the MSCE.

At this point a skeptical reader may be thinking that the correlation seen between recitation attendance and better MSCE score was caused by, or highly influenced by, effort. This is logical but not the case. Figure 5 shows that there are effects of both effort and recitation attendance, but no interaction. [The main effect of recitation attendance thus tutorial use: ($F(4, 201) = 6.184$, $p < 0.001$). The main effect of effort: ($F(1, 201) = 8.480$, $p = 0.004$). No significant interaction: ($F(4, 201) = 0.158$, $p = 0.959$).]

C. *TAs Score Higher on MSCE*

The last comparison presented on construct validity is teaching assistant responses. This data is shown in Figure 6. In all cases the teaching assistants scored at least 15% better (effect size of 3.75) on a pretest than the students were able to on a posttest. In addition, all the TAs did between 20% to 25% better on a posttest (after spending the semester reviewing and teaching the material) than on the pretest.

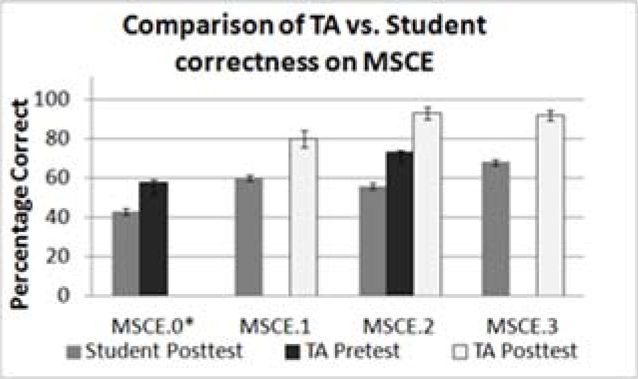


Figure 6: Teaching Assistant MSCE test score.

D. *Psychometric Properties of MSCE.3.*

Reported in Table II are response percentages on the final version of the MSCE test, MSCE.3. The data shown here were collected in a quarter where almost all, 93% on average, students attended recitation with the tutorials. In principle this means that these scores are on the upper end for posttest results. We would expect most course averages to be about 10-20% lower without tutorial use. (A selection of the items is presented later in this paper.)

V. THE ITEMS: AND WHAT THEY INDICATE

A. *Student Understanding of Atomic Structure*

Items #1, 4, 5, 6 (shown on page 6) test different aspects of student understanding of the atomic structure of materials. The questions assess student knowledge of basic terms and definitions like atomic packing factor and body centered cubic. They also assess student knowledge of what the terms mean on atomic scale and look for consistency of student knowledge of the close packed arrangement and what this mean geometrically in terms of the relationship of crystal size to atomic size, which is the most difficult concept for student. Post instruction only about half are able to make the connections. Also, the difference between percent weight and atomic percentage is assessed. A majority of student difficulties on this topic come from a dearth of connections in their knowledge about crystals, atoms, and materials and their low ability to visualize materials on an atomic scale.

B. *Student Difficulties with Material Properties*

Consistent with previous work [1, 20, 21], including our own [11-15], we found that even after instruction students often confound the mechanical properties by either equating them, using them interchangeably, or at very least thinking the properties are necessarily correlated. Thus, students do not see the need for making distinctions between the properties because they think it is unnecessary. This is highlighted by student responses to questions such as: #2 - 28% of students choosing a or b that knowledge of density necessarily implies knowledge of melting temperature; #11 - 34% of students choosing the definition of yield strength when asked for the definition of young's modulus; #12 – 22% reporting that higher yield strength means more strain and load at plastic deformation; #13 – 16% choosing d and e that knowledge of yield strength necessarily implies tensile strength and young modulus (this difficulty was **much** more prevalent in semesters that did not use the tutorials with

Table II: Post Test Statistics on the 31 item MSCE.3, N = 199. Tutorials were used. Correct answers are in bold; this also is the difficulty of each item. Point biserial gives a measure of the item discrimination.

Item	Pt-Bis.	a	b	c	d	e
#1	0.288	5%	14%	8%	**72%**	…
#2	0.300	23%	5%	0%	**72%**	…
#3	0.315	4%	**73%**	10%	13%	…
#4	0.267	7%	6%	**87%**	…	…
#5	0.374	**47%**	23%	8%	22%	…
#6	0.470	**67%**	4%	26%	3%	…
#7	0.404	11%	4%	**65%**	12%	8%
#8	0.238	26%	**66%**	5%	1%	3%
#9	0.443	31%	10%	8%	**39%**	13%
#10	0.410	4%	19%	10%	**67%**	…
#11	0.359	**53%**	7%	6%	34%	…
#12	0.464	3%	**74%**	22%		
#13	0.486	**77%**	1%	6%	8%	8%
#14	0.485	**82%**	17%	1%	…	…
#15	0.297	2%	**94%**	5%	1%	0%
#16	0.318	16%	10%	**74%**	…	…
#17	0.342	8%	3%	6%	**84%**	0%
#18	0.454	22%	3%	**75%**	…	…
#19	0.369	4%	8%	**73%**	16%	…
#20	0.391	27%	**52%**	8%	12%	1%
#21	0.392	28%	14%	0%	**58%**	…
#22	0.369	5%	4%	12%	**79%**	1%
#23	0.371	3%	18%	3%	2%	**75%**
#24	0.399	**66%**	10%	6%	8%	10%
#25	0.337	0%	9%	8%	**83%**	0%
#26	0.431	2%	2%	**75%**	15%	7%
#27	0.261	2%	0%	**93%**	3%	3%
#28	0.376	11%	30%	**57%**	3%	…
#29	0.341	7%	12%	18%	**63%**	…
#30	0.346	5%	**46%**	41%	8%	…
#31	0.410	18%	9%	20%	5%	**47%**

88% of students choosing answers d or e.); #14 - 17% of student choosing the stress/strain graph that goes higher, i.e. has more tensile strength. Also, some students thought that since the higher curve had higher elongation at breaking this was evidence of higher elasticity (again without tutorials we have seen this percentage of incorrect answering be as high as 54%). These examples all highlight the fact that many students conflate the concepts of strength, elasticity, and ductility.

C. Difficulties with Graphs and diagrams

For relatively simple diagrams, such as stress-strain plots (#14) students displayed slope-height confusion, similar to known student difficulties with kinematics graphs in physics [22]. The explanation that "higher position on graphs means more" was commonly found in interviews.

Students also tended to base answers on height rather than slope for concentration vs. position graphs and questions about diffusion. For example, for linear concentration graphs where the points indicated are different heights students will say the higher point has more diffusion even though the slope is the same at both points. For the two questions shown (#8 and #9) students struggle to connect the meaning of the graph to what happens physically with time to the concentration. About a third of students do not understand that material moves down a concentration gradient, and half do not know that a changing flow (and therefor slope) is needed for material to build up. It is clear from students responses and interviews that many students do not understand the meaning of the concentration profiles or their relation to diffusion.

Perhaps as to be expected, students also had great difficulty with novel, unfamiliar graphs and diagrams, such as phase diagrams and isothermal transformation diagrams. Interviews revealed that the difficulty was two-fold. First, the students did not grasp the underlying concepts and content represented in these diagrams. Second, the students were unfamiliar and unpracticed in reading the diagrams and understanding the "rules" of the diagrams.

When considering binary alloys, students often confused the concept of composition and phase fraction. This is particularly true for binary eutectic alloys. In binary eutectic alloys, we found that post instruction, about 25% of students incorrectly believed that the α phase was composed of 100% of one of the metals in the alloy and the β phase was 100% of the other metal. Interestingly, many students who incorrectly believed in "pure phases" still successfully performed lever rule calculations (#20), which inherently assumes that the composition of phases is mixed. In short, many students do not fully understand the nature of the α and β phases as mixtures of elements whose composition may change depending on factors such as temperature.

For TTT diagrams only about half of the students understand how different cooling creates different material strengths and understand that these are not phase or composition differences. In addition, many students cannot read the diagram to know what structure a process gives.

VI. A SELECTION OF MSCE ITEMS

#1. What is the definition of the atomic packing factor?
a. It is the number of atoms in a unit cell.
b. It is the volume of a unit cell that is occupied by atoms.
c. It is how tightly packed atoms are.
d. It is the fraction of a unit cell that is occupied by atoms.

#2. Material A is denser than Material B. How does Material A's melting temperature compare to material B's.
a. Material A has a higher melting temperature than Material B.
b. Material A has a lower melting temperature than Material B.
c. Material A has an equal melting temperature than Material B.
d. a, b, and c are all possible.

#4. What is the close packed direction for body centered cubic crystal?
a. Along the diagonal of a cube face
b. Along the edge of a cube
c. Along a diagonal through the cube center

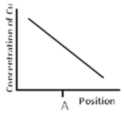
#5. What is the constraining relation between the side length of the unit cell, a, and the radius of an atom, r, for a <u>body centered</u> cubic crystal?
a. $4 * r = \sqrt{a^2 + a^2 + a^2}$
b. $4 * r = \sqrt{a^2 + \sqrt{2} * a^2}$
c. $4 * r = \sqrt{\sqrt{2} * a^2 + 2 * a^2}$
d. $4 * r = \sqrt{a^2 + a^2}$

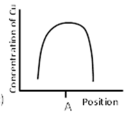
#6. A material is made from 80% wt Mg and 20% wt of Cu. Given that Cu has a greater atomic weight, in grams per mole, than Mg, what is the atomic percent of Cu in the material?
a. less than 20%
b. equal to 20%
c. greater than 20%
d. not enough information

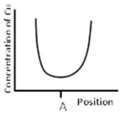
#8 For which of the following concentration vs. position graphs is the net diffusion of copper atoms at point A in the positive x-direction, i.e. to the right as shown?

#9 For which of the following concentration vs. position graphs will the concentration of copper atoms be increasing with <u>time</u> at point A?

a)


b)


c)


d)


(option e is none of the above)

#11. What is the Young's modulus of elasticity or "stiffness" of a material?

a. A measure of a material's resistance to elastic strain when under stress.
b. A measure of a material's ability to return to its original shape after a load is applied.
c. A measure of a material's ability to stretch or deform without breaking.
d. A measure of a material's ability to withstand an applied stress without permanently deforming.

#12. Two pieces of metal A and B are the same size and shape but Metal A has a greater yield strength than Metal B. Which of the following statements is always true?

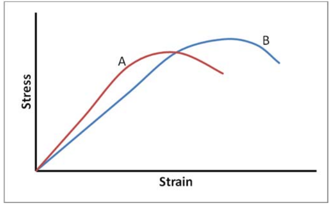
a. Metal A will have greater elastic strain than Metal B for a given applied load.
b. Metal A can withstand a greater applied load than Metal B before permanent deformation occurs.
c. Both a and b are true

#13. Two pieces of metal, A and B, are the same size and shape but Metal A has a greater yield strength than Metal B. Which of the following statement is always true?

a. Metal A will permanently deform at a greater stress than Metal B
b. Metal A will have a greater tensile strength than Metal B
c. Metal A will have a greater young's modulus of elasticity than Metal B
d. Both a & b
e. a, b, & c are all true

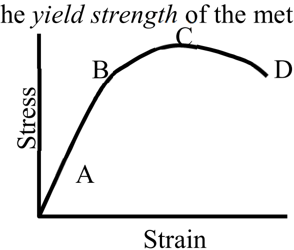
#14. Which metal has a higher modulus of elasticity?

a. A has a higher modulus.
b. B has a higher modulus
c. The modulus of A is equal to that of B

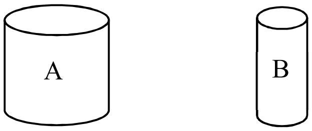


#15. To the right is a stress-strain curve for a metal. Which part of the curve defines the *yield strength* of the metal?

a. A
b. B
c. C
d. D
e. All of the above



#16. The following metal pieces are cut from the same plate. Compare the yield strength of the pieces.



(A and B have equal heights)

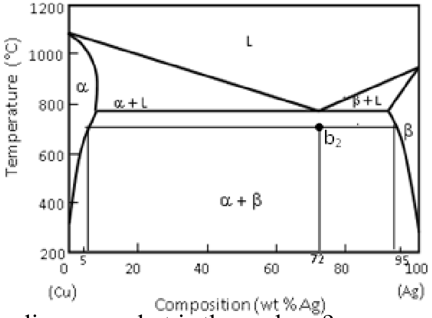
a. A has a higher yield strength than B
b. B has a higher yield strength than A
c. A and B have the same yield strength

#19. Metal A has a higher yield strength than metal B. Which of the following is the best statement describing the ductility of the metals?

a. Metal A will be more ductile
b. Metal B will be more ductile
c. Both a) and b) are possible
d. Ductility depends only on tensile strength, rather than yield strength

#20. At point b2 what is the <u>fraction of the alloy</u> that is α?

a. 6%
b. 26%
c. 28%
d. 72%
e. 94%

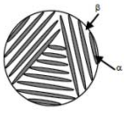


#21. In the phase diagram, what is the α phase?

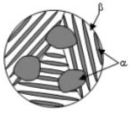
a. It is only Copper Atoms.
b. It is only Silver Atoms.
c. It is a mixture of Copper and Silver atoms with a specific fixed solubility.
d. It is a mixture of Copper and Silver atoms with a solubility that depends on temperature.

#22. Which of the following sketches best illustrates the microstructure that will develop when a 85 wt% Ag alloy is cooled slowly from 1000°C to 250°C? [Refer to the phase diagram in #20/21]

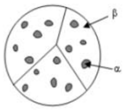
#23. Which of the following sketches best illustrates the microstructure that will develop when a 6 wt% Ag alloy is cooled slowly from 1200°C to 250°C? [Refer to the phase diagram in #20/21]



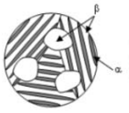
A



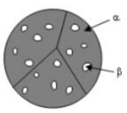
B



C



D



E

#29. Refer to the TTT diagram [shown below]. Austenite at the eutectoid composition is quenched from 750°C to 625°C and held for 10 s. It is then quenched to 300°C and held for 10 s. Finally, it is quenched to room temperature (25°C). What is the final microstructure?

a. 100% austenite
b. 100% martensite
c. 50% pearlite + 50% austenite
d. 50% pearlite + 50% martensite

#30. Which of the following will always be effective in distinguishing a difference between a sample of austenite and a sample of pearlite?

a. The weight percent of Carbon is different between the two materials
b. The phase(s) present in the two materials are different
c. Both a & b
d. There are no composition or phase differences between the two materials

#31. Which of the following is the correct ranking of the metals, from greatest to least strength, for a fixed wt % of carbon? (**Isothermal transformation diagram for Iron-Carbon alloy at eutectoid composition was provided for the students. This was Figure 10.13 from the Calister text [23])

a. Martensite, Austenite, Pearlite
b. Austenite, Martensite, Pearlite
c. Pearlite, Austeniste, Martensite
d. Pearlite, Martensite, Austenite
e. Martensite, Pearlite, Austenite

A. MSCE Items not Shown Here

The following items were not shown here: #3 on melting temp; #7 on isotropy; #10 on diffusion; #17, #24 on defects and strength; #18 on applied mechanical properties; #25 on single phase transitions; #26 on fatigue; #27 and #28 on creep. While these items do tell us interesting things about student ideas and reasoning there is insufficient space to present them in this article.

B. Comparison of MSCE vs. MCI Item Types

The MCI and the MSCE have many questions that are not covered on the other test (see Table III). This makes the two tests somewhat complimentary to each other. For example, the MCI has many more questions about the basic chemical and physical rules of materials like: how atoms in solids can move, what phases Ni can exhibit, why glass is clear, what causes electrical conductance, "colder" is a property of thermal conductivity, and a few others. The MSCE does not have these types of questions but instead asks many questions about crystal structure, atomic packing factor, and geometric conceptions of crystals. The MCI has several questions on polymers which the MSCE does not cover. But, the MSCE has many questions on phase diagrams and TTT plots which the MCI does not cover. We envision these two tests being used in compliment to each other to provide instructors with valuable information on their students' understanding and the effects of their instruction.

Table III: Different Concepts Covered by MCI and MSCE

Topics	# of items on MCI	# of items on MSCE
Chemical and Physical Properties	9	1
Diffusion	1	3
Geometry	2 (of a Cube)	2 (BCC crystal) 1 isotropy 1 def. of APF
Mechanical Properties	2 (Melting Temp. and Bonding)	9 (Yield Strength, Tensile Strength, Elasticity/Strain, Ductility, Young's Modulus)
Conductivity	2	0
Defects and Strength	4	2
Applied Mech. Prop.	3	1
Failure	2 (Fracture)	3 (Creep/Fatigue)
Polymers and Composites	4	0
Phase and TTT Diagrams	0	5 (Phase) 3 (TTT)
Other	1	0

VII. CONCLUSIONS

Here we reported on the creation and validation of a Materials Science Conceptual Evaluation (MSCE). This assessment was created from a set of test items designed to identify areas of difficulties students had with the introductory materials science course content. Thus, a majority of items were separately validated before inclusion in the MSCE test version.

The tests items were created by the following process: 1. identify a difficulty through interviews, literature searches, and old test questions; 2. create a multiple-choice item from short-answer and interview responses; 3. establish validity.

The content validity of the test items was established by three course instructors, and materials scientists, assessing items based on importance and correctness of the answer options and question. This data was analyzed and items with low importance were not included on the MSCE. In addition, items were reworded as needed to make them technically sound and clearer based on instructor feedback. The construct validity was established through a series of think-aloud interviews and explain-your-answer short responses while the student was taking the MSCE test. Finally, a series of statistical validation measures were taken to ensure the items were valid and the test was internally reliable. There are several statistical measures reported here to ensure the validity of the MSCE. Correlations with final exam score, point biserial coefficients, reliability of the test as a whole (KR 20), gains with increased tutorial participation and effort on the test, and better performance by more expert students, i.e. the TAs.

These results lend strong support to the claim that the MSCE is a valid and appropriate measure of student conceptual understanding for an introductory materials science course. The next step for the MSCE test is to collect data from other courses and institutions. This would provide information on the validity of the test for a wider range of students and introductory course formats. Also, we would like to be able to run IRT and Rausch analysis for the test, but these require many more students than can be collected at one institution alone. We hope this article will garner volunteers so that this data can be collected.

ACKNOWLEDGMENT

Special thanks to Katharine Flores the instructor for the materials science course. Her knowledge of student misconceptions in materials science and innovative teaching were in valuable for this project. Funding for this research was provided by the Center for Emergent Materials: an NSF MRSEC under award number DMR-1420451.

REFERENCES

[1] Krause, S., Decker, J. C., Niska, J., and Alford, T., *A Materials Concept Inventory for introductory materials engineering courses*, National Educators Workshop Update 2002, **17**, 1-8 (2002).

[2] Beichner, R. J., *Testing student interpretation of kinematics graphs*. Am. J. Phys., **62**, 750-762, (1994).

[3] Engelhardt, P. V., *An Introduction to Classical Test Theory as Applied to Conceptual Multiple-choice Tests.* In Getting Started in PER, edited by C. Henderson and K. A. Harper. (AAPT, College Park, MD, 2009).

[4] NGSS Lead State. 2013. *Next Generation Science Standards: For States, By States.* Washington, DC. The Nations Academies Press.

[5] Hestenes, D., Wells, M., and Swackhamer, G. *Force concept inventory*. The Physics Teacher, **30**, 141-166 (1992).

[6] Lawson, A. E., *The development and validation of a classroom test of formal reasoning.* Journal of Research in Science Teaching, **15**, 11-24 (1978).

[7] Rebello, S. N. and Zollman, D. A. *The effect of distracters on student performance on the force concept inventory.* Am. J. Phys., **72**, 116-125 (2004).

[8] Lindell, R., Peak, E., and Foster, T. M. Are they all created equal? A comparison of different concept inventory development methodologies. Physics Education Research Conference **883**, 14-17 (2007).

[9] Wang, J. and Bao, L, *Analyzing Force Concept Inventory with Item Response Theory.* American Journal of Physics, **78**, 1064-1070 (2010).

[10] Rosenblatt, R., Heckler, A. F., & Flores, K. "A Tutorial Design Process Applied to an Introductory Materials Engineering Course." *Advancements in Engineering Education,* 2013.

[11] Rosenblatt, R., Heckler, A. F., & Flores, K. "A Tutorial Design Process Applied to an Introductory Materials Engineering Course." *Advancements in Engineering Education,* 2013.

[12] Rosenblatt, Rebecca. "Identifying and addressing student difficulties and misconceptions: examples from physics and from materials science and engineering." *Electronic Thesis or Dissertation. Ohio State University*, 2012.

[13] Rosenblatt, R. J. and Heckler, A. F. "Student Understanding of the Mechanical Properties of Metals in an Introductory Materials Science Engineering Course", ASEE Conference, 2010.

[14] Heckler, A, F, and Rosenblatt, R, J, "Student Understanding of Atomic Bonds and their Relation to Mechanical Properties of Metals in an Introductory Materials Science Engineering Course", ASEE Conference, 2010.

[15] Heckler, A, F, and Rosenblatt, R, J, "Student Difficulties with Basic Concepts in Introductory Materials Science Engineering", FIE Conference, 2011.

[16] Ding, L., Chabay, R., Sherwood, B., and Beichner, R., *Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment.* PRST-PER, **2**, 010105 (2006).

[17] Thornton, R. K. and Sokoloff, D. R. Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. American Journal of Physics, **66**, 338-352 (1998).

[18] Messick, S., Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, **50**, 741-749 (1995).

[19] Streveler, R. A., Litzinger, T. A., Miller, R. L. and Steif, P. S., Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions. JEE, 97: 279–294 (2008).

[20] Kitto, K.L., "Analyzing What Students Write about Materials - Another Strategy for Developing Conceptual Knowledge in a Materials Engineering Course." ASEE/IEEE Frontiers in Education Conference, S2G-14-8, 2007.

[21] Kitto, K. L., "Developing and Assessing Conceptual Understanding in Materials Engineering Using Written Research Papers and Oral Poster Presentations", *38rd ASEE/IEEE Frontiers in Education Conference Proceedings,*F4A-1-6, 2008.

[22] McDermott, L, C, Shaffer, P, S, and the Physics Education Group at the U. of Washington, *Tutorials in Introductory Physics*, Prentice Hall, 2002.

[23] Callister, W, D, *Materials Science and Engineering: an Introduction*, Wiley, New York, 2007.