A Tale of Four Questions: Unexpected Insights from Think-Aloud Interviews

Debra Dillner[a], Shirley Lin[a], Maria Schroeder[a], Melonie Teichert[a], Diane Bunce[b], and JudithAnn Hartman[a]

[a] Chemistry Department, USNA
[b] Department of Chemistry, The Catholic University of America

USNA Conference on Teaching and Learning
15 MAY 2018

Session Goals

Participants will be able to…

- Explain how information about student learning can be gathered from multiple choice questions used on course-wide exams
- Describe how think-aloud interviews on solving multiple choice questions can give insight into student learning
- Identify the differences in problem-solving behaviors observed by a variety of students
- Recognize opportunities in their own courses to apply these techniques to improve student learning

Poll: Use of Multiple Choice Questions

Plebe Chemistry Practices:

Common exams

- 6-week (25 MC, 10%)
- 12-week (25 MC, 10%)
- Final (70-80 MC, 25%)

Instructor-written instruments (30%)

- Variety of formats (not all MC)

Exam Question Database

- For more than 20 years, the Chemistry Department has maintained a database of exam questions used on common exams

- Allows for better assessment
  - Reuse questions periodically
  - Track performance over time

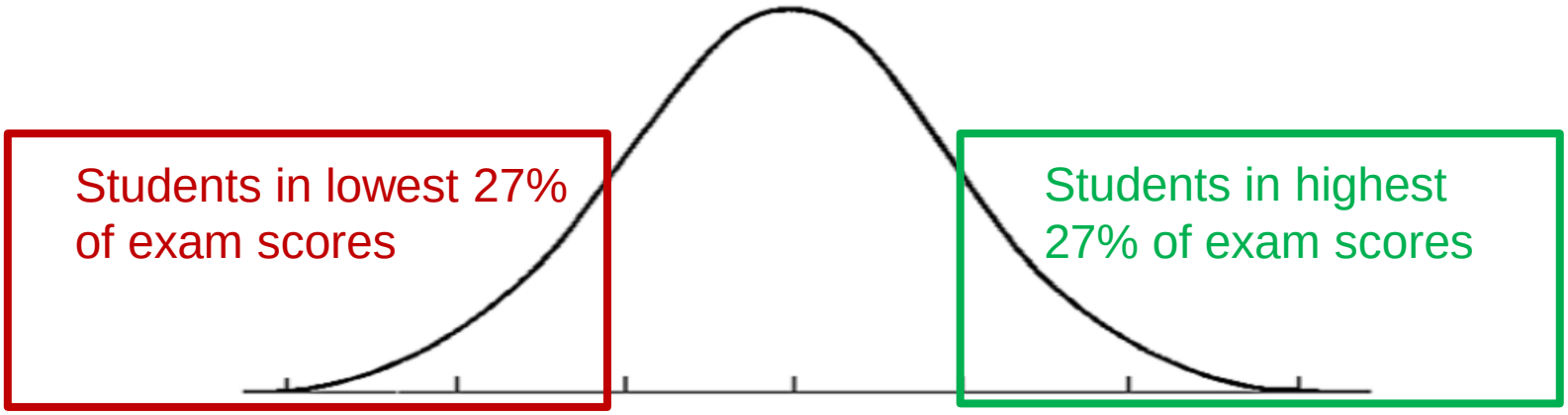
- All 4 interview questions were selected from the exam database

- Each question recorded includes dates of use and **item statistics**

Item Statistics

Item difficulty = % students choosing correct answer

Discrimination index



For each question, tabulate answers for each group

Discrimination index = Fraction of students in highest group choosing correct answer − Fraction of students in lowest group choosing correct answer

Item Difficulty and Discrimination Index

Interview questions represented a range of problem-solving skills (algorithmic, conceptual)

	Question #			
	1	2	3	4
Item Difficulty (%)	65	80	69	87
Discrimination Index	0.49	0.33	0.42	0.30

Discrimination index above 0.20 is considered discriminating

A question with a very high or very low item difficulty does not discriminate well

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). Measurement and evaluation in psychology and education (5th Ed.). New York: MacMillan.

Towns, M. H. “Guide to Developing High-Quality, Reliable, and Valid Multiple Choice Assessments.” *J. Chem. Educ.* **2014***91*, 1426−1431.

Examining Individual Multiple Choice Questions

- Item Difficulty and Discrimination Index are simple statistics that can provide some information about student learning

- Choice of distractor can give a more detailed picture of student understanding and misunderstanding

Interview Question 1

Calculate the value of ΔH°_{rxn} for the reaction

$$Mg\,(s) + 2\,HCl\,(aq) \rightarrow MgCl_2\,(aq) + H_2\,(g)$$

Substance	ΔH°_f (kJ/mol)
HCl (aq)	-167.2
$MgCl_2$ (aq)	-641.6

a. -307.2 kJ ← **CORRECT ANSWER**

b. -473.8 kJ

c. -976.0 kJ

d. Not enough information is given

DISTRACTORS

Examining Distractors: Question 1

Groups by Exam Score	Item Difficulty (%)	Most Popular Distractor (% students choosing)
Total Population	65	d. (23%)
Score ≥ 88%	88	d. (8%)
Score 72-84%	64	d. (22%)
Score ≤ 68%	39	d. (41%)

Why was “d.” the most popular distractor?

Interview Question 1

Calculate the value of ΔH°_{rxn} for the reaction

$$Mg\,(s) + 2\,HCl\,(aq) \rightarrow MgCl_2\,(aq) + H_2\,(g)$$

Substance	ΔH°_f (kJ/mol)
HCl (aq)	-167.2
$MgCl_2$ (aq)	-641.6

a. -307.2 kJ
b. -473.8 kJ
c. -976.0 kJ
d. Not enough information is given

Didn't recall that elements have ΔH°_f = 0 kJ/mol

Discussion

- How do you analyze multiple choice question performance in your courses?

Methodology for Interviews

IRB APPROVAL # USNA.2015.0001-CR03-EP7-A

Recruitment

- Students invited during lab periods (50%) and Academic Center supplemental instruction courses (50%)

- Students who were experiencing trouble in course were encouraged to participate

- Interviews conducted by current or emeritus USNA faculty members

Methodology for Interviews

Interviews

- Students asked to solve problems aloud (“think-aloud”)
    - Minimum of questioning from interviewer

- 4 multiple-choice questions selected from 12-week common exam
    - Two algorithmic (#1 and 4)
    - Two conceptual (#2 and 3)
    - Completed within previous 5 days of exam

- Answer keys not yet released

- Interviews recorded and transcribed for coding

Demographics of Interviewees

- 72 students (Plebes) interviewed ranged in ability, with more lower-ability students

SC111 Class (w/o interviewees) Final Course Grades	*N*	%
A	166	18.0
B	341	37.1
C	291	31.6
D	111	12.1
F	11	1.2
Total	920	

SC111 Interviewees Final Course Grades	*N*	%
A	14	19.4
B	13	18.1
C	23	31.9
D	20	27.8
F	2	2.8
Total	72	

- Interview pool mimics the overall population but is statistically different based on performance

Grouping Students by Ability

- Based on the average exam score (6-week, 12-week, and final common exams), interview students were placed into ability groups

Ability	***N* (%)**	**Avg Exam Score (%)**
Top	19 (26%)	≥ 85
Middle	20 (28%)	70 - 84.99
Bottom	33 (46%)	< 70
Total	72	

Interview Question 1 (Mostly Algorithmic)

Calculate the value of ΔH°_{rxn} for the reaction

$$Mg\,(s) + 2\,HCl\,(aq) \rightarrow MgCl_2\,(aq) + H_2\,(g)$$

Substance	ΔH°_f (kJ/mol)
HCl (aq)	-167.2
$MgCl_2$ (aq)	-641.6

a. -307.2 kJ ← **CORRECT ANSWER**

b. -473.8 kJ

c. -976.0 kJ

d. Not enough information is given

} **DISTRACTORS**

Question 1 Interview Quotes

"I know that if it's a **diatomic molecule** like if it's H_2 a gas I know that **it equals 0 for its enthalpy**. And, also **magnesium equals 0 in its natural state**."

“I crossed out magnesium in its solid state because… **solids don’t have enthalpies of reaction**. **Hydrogen gas, that’s also in its standard state**.”

".... umm **I don't know how you would find the delta H's of things that aren't given to you** ..."

Question 2 (Mostly Conceptual)

Choose the correct statements about the reaction below:

$$C_2H_5OH\ (l) + 3\ O_2\ (g) \rightarrow 2\ CO_2\ (g) + 3\ H_2O\ (l)$$

$$\Delta H = -1.37 \times 10^3\ kJ$$

I. The reaction is endothermic.
II. The reaction is exothermic.
III. The enthalpy value would be different if the H_2O formed was gaseous.
IV. The reverse reaction would have the same value of ΔH.
V. Using 2 moles of C_2H_5OH (l) would result in a ΔH value of -6.85×10^2 kJ.

a. II, III and V
b. II and III ← **CORRECT ANSWER**
c. II and IV
d. II and V

DISTRACTORS

Question 2 Interview Quotes (Test-Taking Strategy)

"The delta H is negative. And I think that means its exothermic? **So well it has to be because [statement] II is in all the answers**…"

" I guess I didn't exactly know everything but **I just used process of elimination on a lot of them and...just looked at the answer choices**"

Question 2 Interview Quotes (Language)

$$C_2H_5OH\,(\ell) + 3\,O_2\,(g) \rightarrow 2\,CO_2\,(g) + 3\,H_2O\,(\ell)$$

"**C_2H_5OH liquid** plus 3 **oxygen gas** yields 2 **carbon dioxide gas** plus 3 **liquid water**..."

"**C_2H_5OH** plus 3 **O_2** yields 2 **CO_2** plus 3 **H_2O**."

Question 3 (Conceptual)

Using the table below, determine which test followed by a physical separation (centrifuge and decant supernatant) can be used to separate Cu^{2+} ions from a solution that also contains Cr^{3+} and Fe^{3+} ions.

Metal ion	**Test 1 Conc. NH_3 reagent**	**Test 2 Basic H_2O_2 reagent**	**Test 3 KSCN reagent**
Cr^{3+}	Form gray-green solid $Cr(OH)_3$	Form yellow solution of CrO_4^{2-}	No reaction
Cu^{2+}	Form deep blue solution of $Cu(NH_3)_4^{2+}$	Form pale blue solid $Cu(OH)_2$	No reaction
Fe^{3+}	Form rust red solid $Fe(OH)_3$	Form rust red solid $Fe(OH)_3$	Form blood red solution of $FeSCN^{2+}$

a. test 1 ← **CORRECT ANSWER**

b. test 2

c. test 3

d. none of these tests

DISTRACTORS

Question 3 Interview Quotes (Conceptual)

"**So, this one I'm looking for a test that separates not just $Cr^{3+}$and Fe^{3+} would be either a precipitate or a supernatant and then Cu^{2+} would be the opposite.** ...So A, test one. **Let me just make sure the rest are [incorrect]**… "

"My mindset, I really didn't understand the question too well but the way I think I got the right answer was…. **I saw that test 3 contained Fe let's see yeah Fe^{3+} or Fe so I was kinda just like well I'm going to guess and put it in the middle so I just selected test 2.** I really didn't know how to go about that problem so I looked for any kind of clue possible."

Question 4 (Algorithmic)

What is the energy of a 405 nm photon of light?

a. 4.91×10^{-19} J ← **CORRECT ANSWER**

b. 4.91×10^{-9} J

c. 8.96×10^{-40} J

d. 8.96×10^{-49} J

e. 2.69×10^{-40} J

DISTRACTORS

Question 4 Interview Quotes (Language - Units)

"We're given wavelength which is **405 nanometers**, we know c cause that's given just on our sheet which is **2.9979×10^{8} meters per second** and we also know **Planck's constant** which is **6.626×10^{-34} Joules per second**. **So, umm right off the front, we can see that if wavelength is in nanometers but c is in meters so we need to change nanometers to meters."**

"So I think it was, my mindset was since I saw the question was asking for **405** I was just **for some reason every answer option that started with a four I was like well it's probably not that one** so I just kinda crossed off B and C. And then I was like well it's It had to be A, D and E because those are all different numbers that were farther away from **405** so I believe I chose … **2.69**."

Research Approach

- **Thematic Analysis**
  - Related to **grounded theory** and **phenomenology**
    - Data-driven, inductive process focusing on semantic themes
    - Identifying and analyzing themes through **coding**
    - Rigorous, systematic way of tracking codes
- Can provide a method of translating **qualitative** data into **quantitative**

Glaser, B.; Strauss, A. The Discovery of Grounded Theory. Aldine: Chicago, 1967.

Corbin, J.; Strauss, A. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qual. Soc.* 1990, 13 (1), 3-21.

Patton, M. Q. Qualitative Research and Evaluation Methods, 3rd edition. Sage Publications: Thousand Oaks, 2001.

Quantitative Analysis of Interviews: Coding

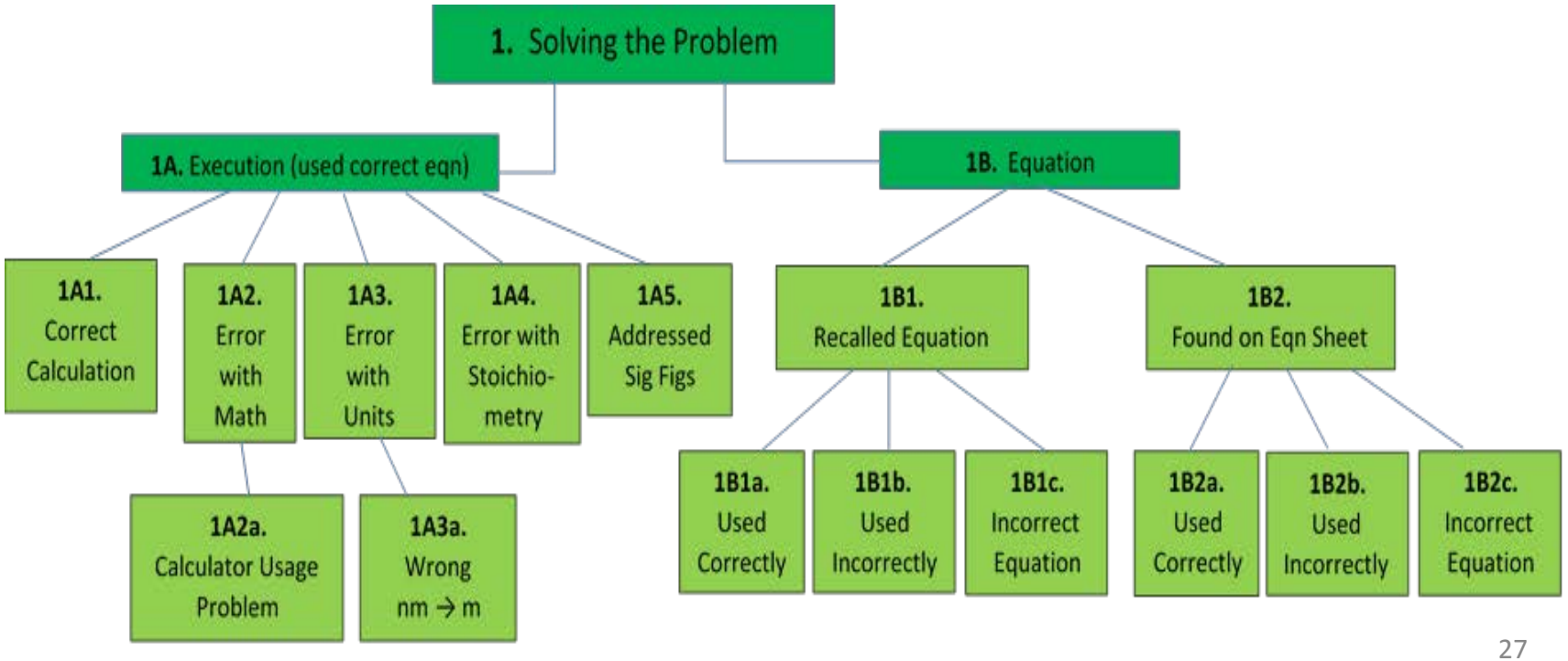
Main codes in 4 categories:

- Problem-solving
- Concept understanding
- Use of language
- Test-taking strategies

Quantitative Analysis of Interviews: Coding

Main codes in 4 categories:

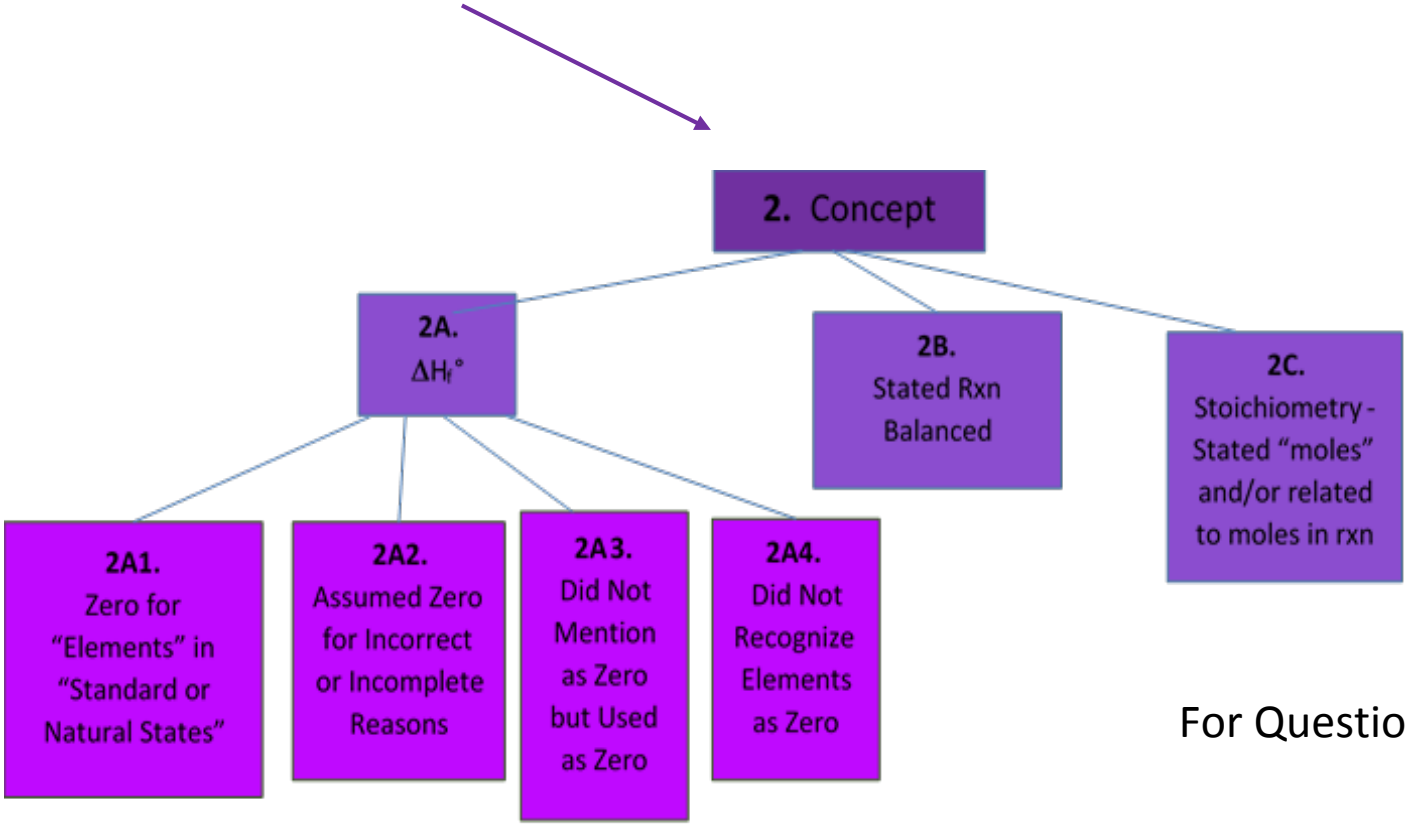
- Problem-solving



Quantitative Analysis of Interviews: Coding

Main codes in 4 categories:

- Problem-solving
- Concept understanding

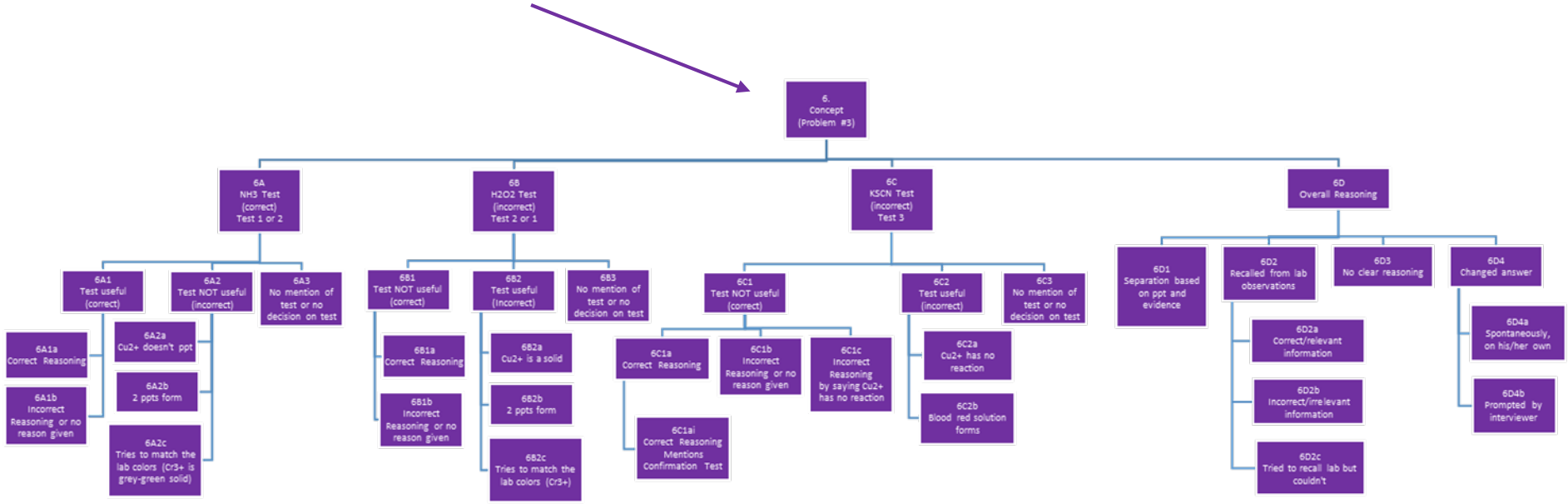


For Question 1

Quantitative Analysis of Interviews: Coding

Main codes in 4 categories:

- Problem-solving
- Concept understanding

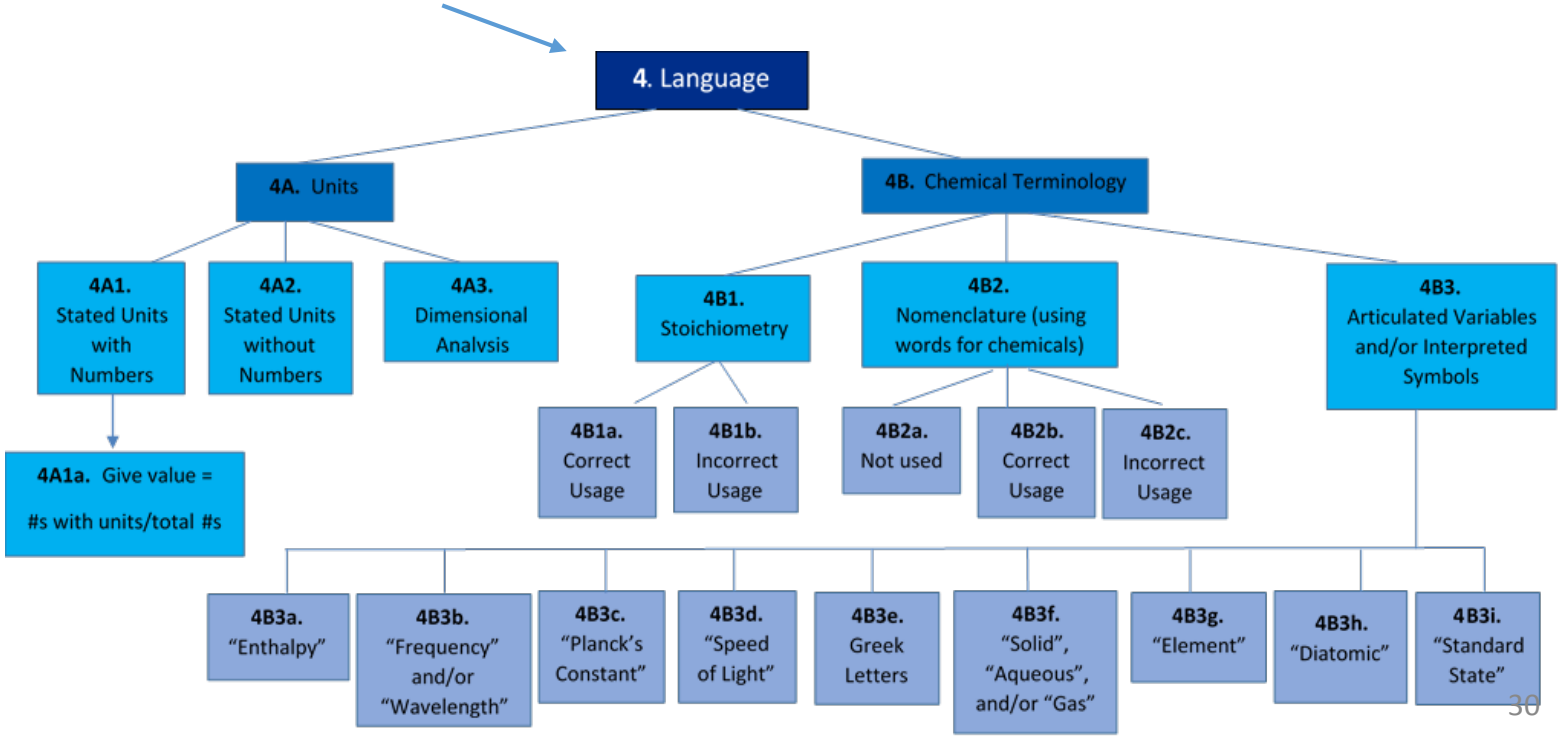


For Question 3

Quantitative Analysis of Interviews: Coding

Main codes in 4 categories:

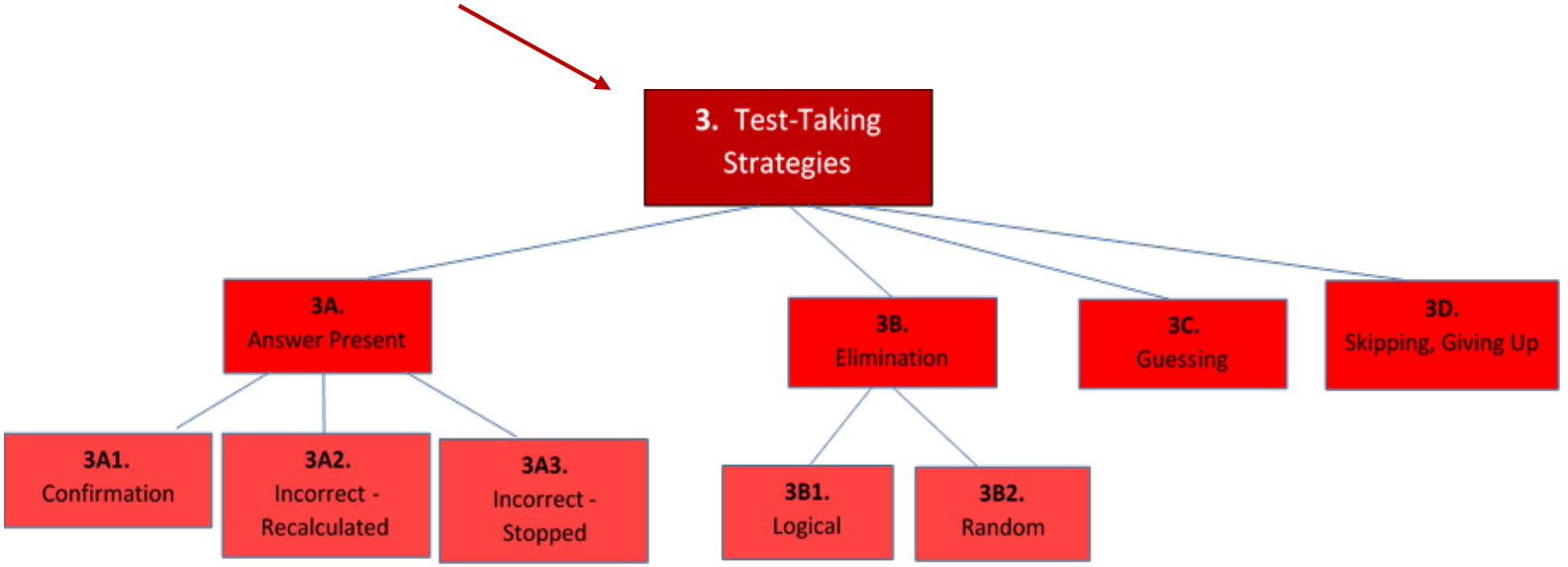
- Problem-solving
- Concept understanding
- Use of language



Quantitative Analysis of Interviews: Coding

Main codes in 4 categories:

- Problem-solving
- Concept understanding
- Use of language
- Test-taking strategies



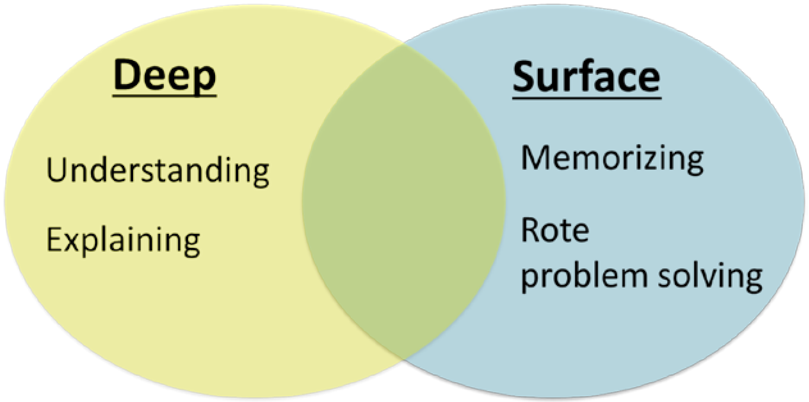
Deep vs. Surface Learners and Behaviors

- Deep Learners

  Intrinsically motivated (satisfaction of learning)

- Surface Learners

  Extrinsically motivated (grades)

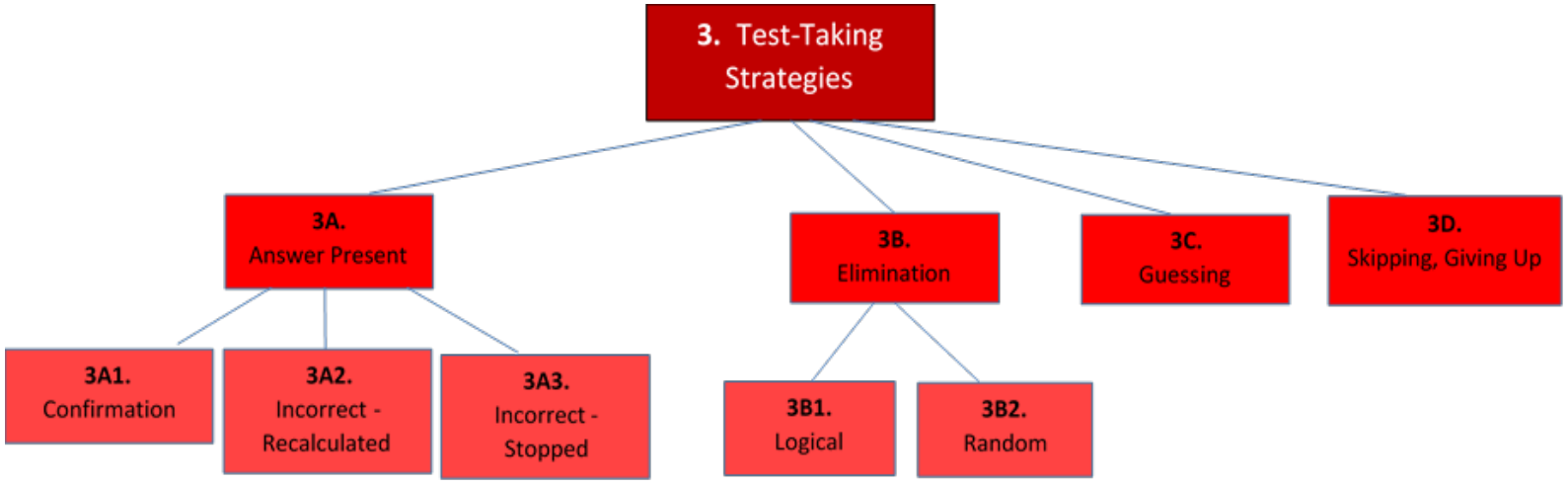


Chin, C.; Brown, D. E. Learning in Science: A Comparison of Deep and Surface Approaches. *J. Res. Sci. Teach.* 2000, 37 (2), 109−138.

Sinapuelas, M. L. S.; Stacy, A. M. The Relationship Between Student Success in Introductory University Chemistry and Approaches to Learning Outside of the Classroom. *J. Res. Sci. Teach*. 2015, 52 (6), 790−815.

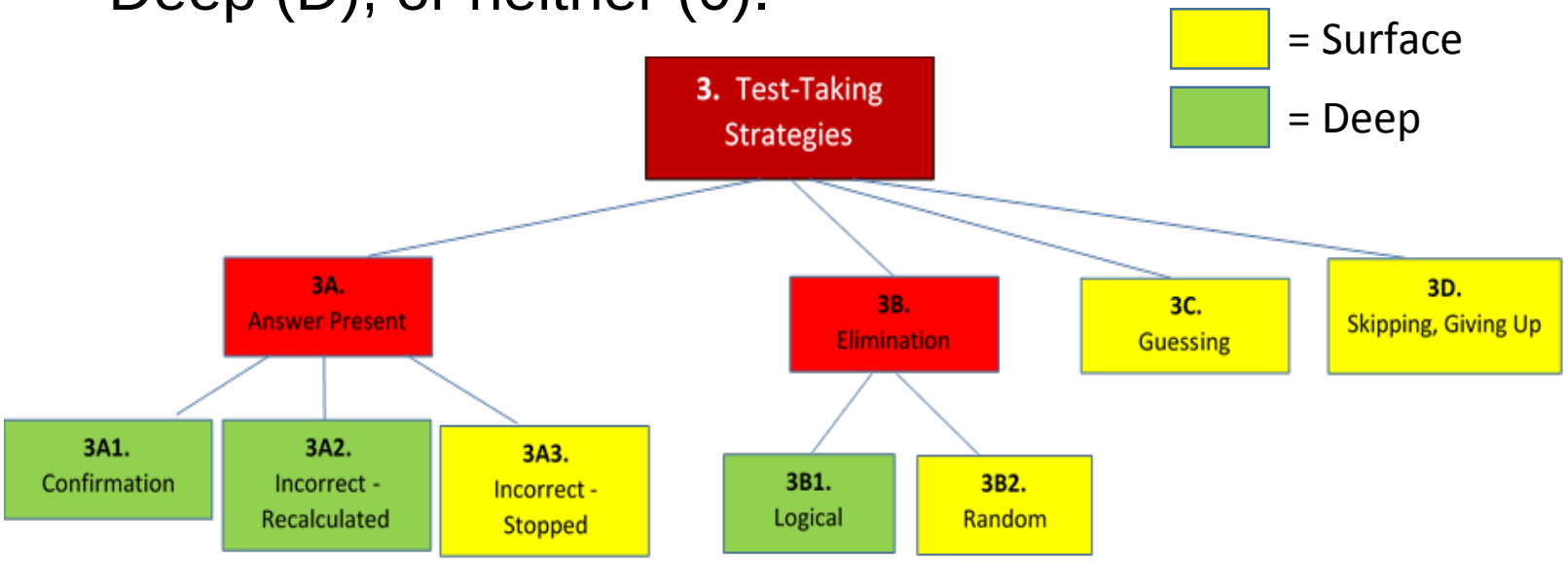
Coding Analysis: Test-taking Code Example

- After developing and testing the coding scheme, we coded all 4 questions in each of the 4 coding categories
- Codes were recorded in a spreadsheet
- We determined which codes were Surface (S), Deep (D), or neither (0).



Coding Analysis: Test-taking Code Example

- After developing and testing the coding scheme, we coded all 4 questions in each of the 4 coding categories
- Codes were recorded in a spreadsheet
- We determined which codes were Surface (S), Deep (D), or neither (0).



Coding Analysis: Test-taking Code Example

- For each student, S and D totals for each problem and coding category were determined and normalized:

$$\%D = \frac{total\ Deep}{total\ \max Deep} \text{ x } 100$$

$$\%S = \frac{total\ Surface}{total\ \max Surface} \text{ x } 100$$

$$\%\boldsymbol{D} - \%\boldsymbol{S} = \textbf{Global code score}$$

Lower
%D-%S

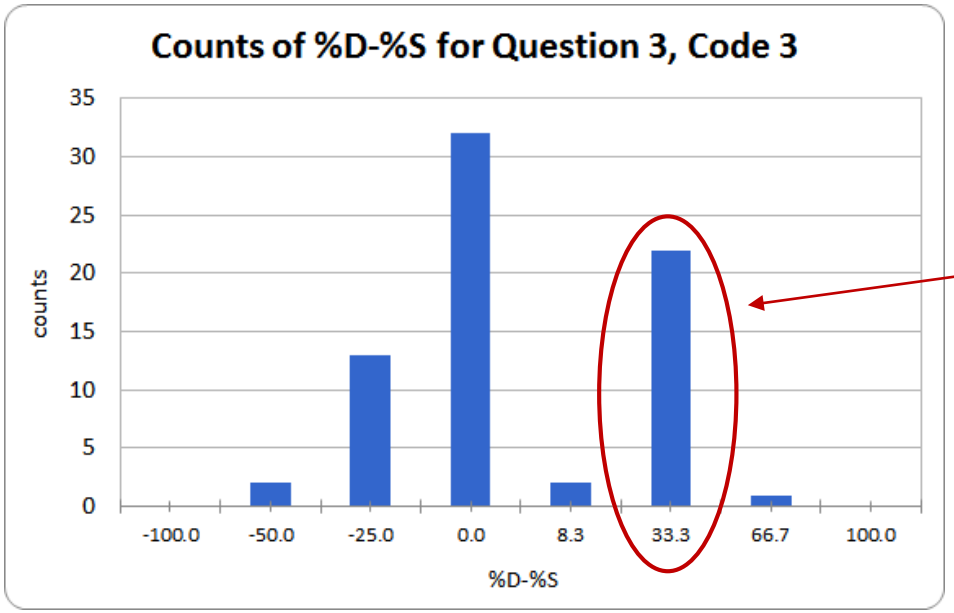
Surface Approach

Higher
%D-%S

Deeper Approach

Coding Analysis: Test-taking Code Example

- For Question 3, code Test-taking Strategies:



Who are these students who showed deeper approaches?

What did they do?

Lower %D-%S

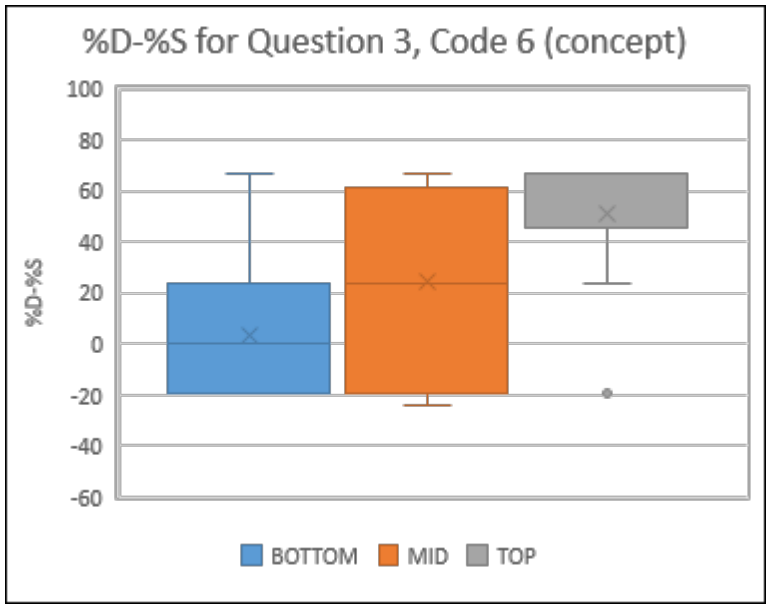
Surface Approach

Higher %D-%S

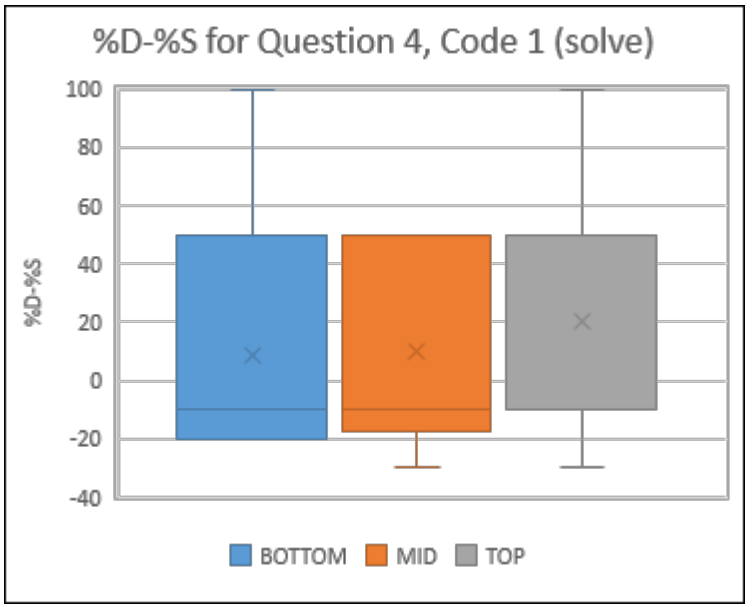
Deeper Approach

Coding Analysis: Some Preliminary Results

- We have a lot of data – could be analyzed many ways



All 3 ability groups are statistically different ($p < 0.05$)

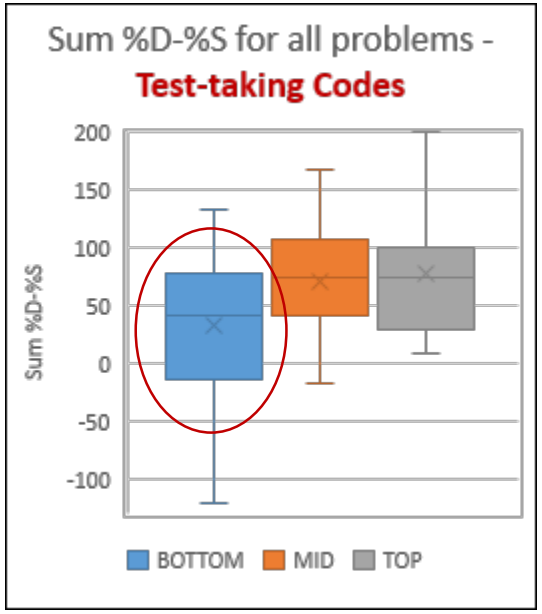


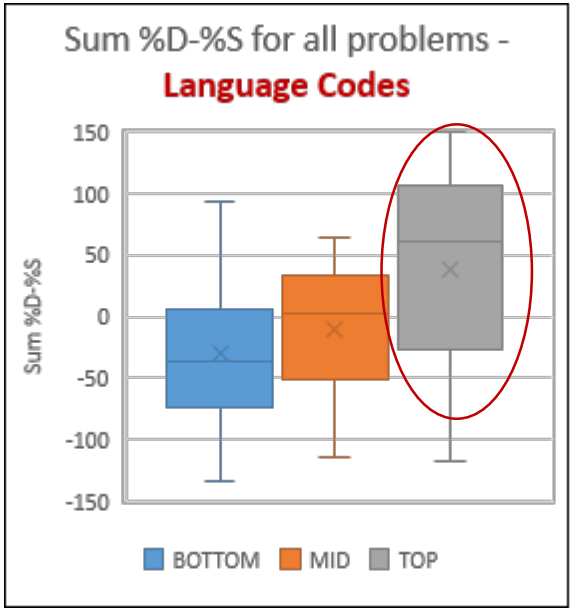
The 3 ability groups are not statistically different ($p > 0.05$)

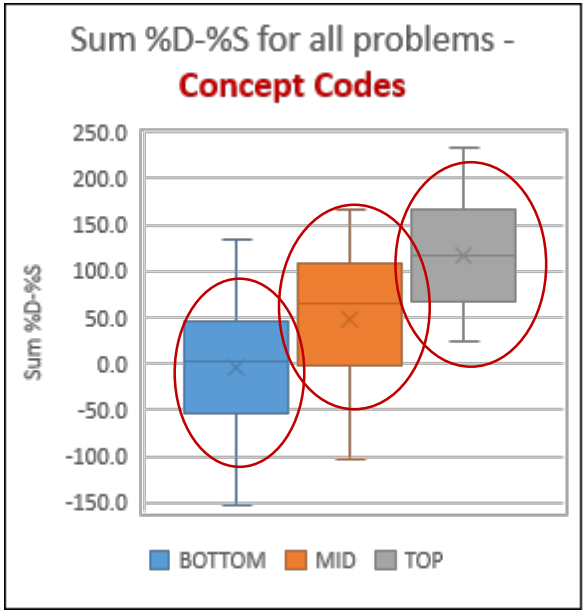
- Question 3 (conceptual) may provide more information about surface/deep approaches than Question 4 (algorithmic).

Coding Analysis: Some Preliminary Results

- Codes can be combined across all 4 questions







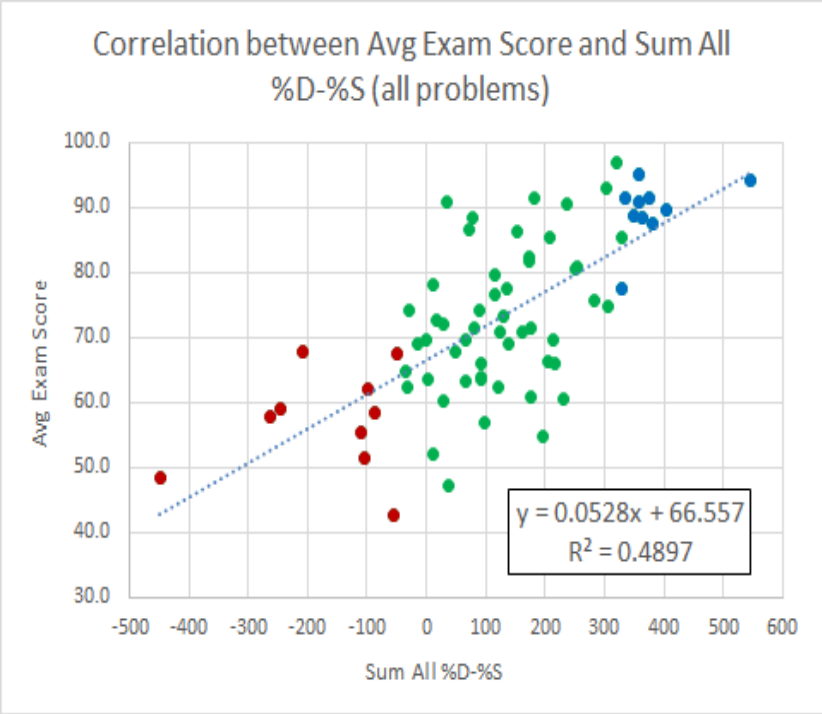
Test-taking strategies: Bottom students different than Middle or Top

Language: Top students different than Middle or Bottom

Concept: All 3 groups are different from each other

Coding Analysis and Relationship to Performance

Lowest 10 "Surface" Students		
Sum All %D-%S	Avg Exam Score	Final Grade
-50.7	67.7	C
-55.8	42.7	F
-86.4	58.5	D
-99.5	62.2	C
-103.5	51.3	D
-110.3	55.5	D
-207.9	67.8	C
-244.7	58.9	D
-263.9	57.9	D
-448.2	48.5	D



Highest 10 "Deep" Students		
Sum All %D-%S	Avg Exam Score	Final Grade
330.1	77.4	B
335.3	91.5	A
348.6	88.7	A
356.9	95.2	A
357.6	91.0	A
363.1	88.4	A
375.8	91.3	A
380.2	87.5	B
403.6	89.7	B
545.0	94.3	A

Our coding analysis seems to correlate to performance for high and low achieving students, but what about the middle students?

Summary and Implications

- Performance on multiple choice questions provides limited insight into student understanding
- Think-aloud interviews reveal:
  - further insight into the depth of student understanding
  - effect of question format on the use of surface approaches
  - differing use of deep vs. surface approaches related to achievement level
- Incorporating a variety of question formats is necessary to reveal the full extent of student understanding

Discussion

Brainstorm ways in which we as instructors can encourage practices associated with deep learning rather than surface learning.

Acknowledgements

Mr. James Kinnear

USNA Chemistry Department

Midshipmen enrolled in SC111 Fall 2014

1st Lt Kate Smith and LTJG Jordan Snyder

USNA IRB Office