

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

DBER Speaker Series

Discipline-Based Education Research Group

10-2013

Question Order Effects on a General Chemistry Concept Inventory

Travis Lund

University of Nebraska-Lincoln, travis.lund@oit.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/dberspeakers>



Part of the [Curriculum and Instruction Commons](#), [Educational Methods Commons](#), [Higher Education Commons](#), and the [Science and Mathematics Education Commons](#)

Lund, Travis, "Question Order Effects on a General Chemistry Concept Inventory" (2013). *DBER Speaker Series*. 44.

<https://digitalcommons.unl.edu/dberspeakers/44>

This Presentation is brought to you for free and open access by the Discipline-Based Education Research Group at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in DBER Speaker Series by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Abstract for DBER Group Discussion on 2013-10-24

Presenter(s), Department(s):

Travis Lund
Postdoctoral Research Associate
Department of Chemistry
University of Nebraska-Lincoln

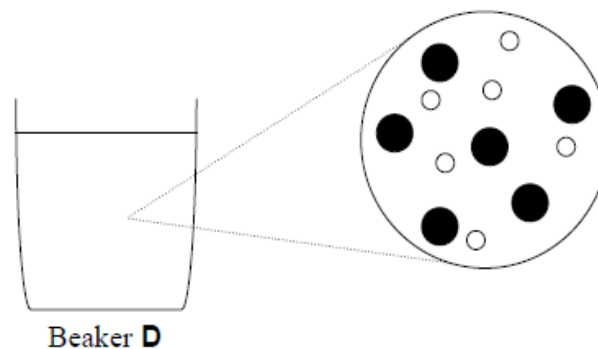
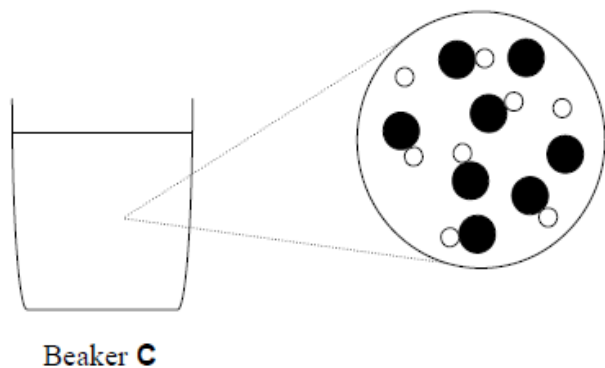
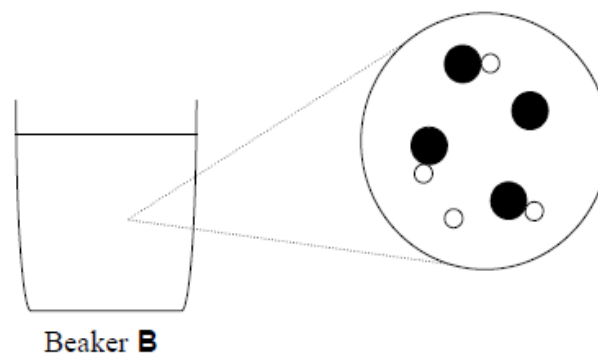
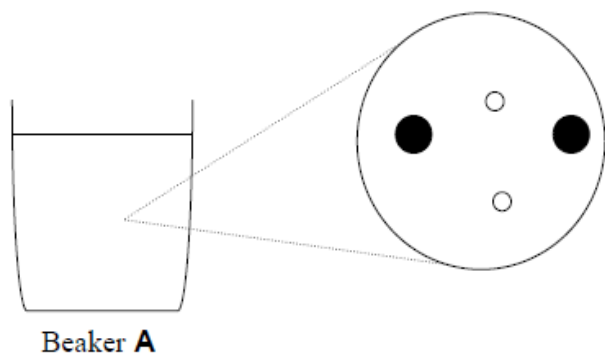
Title:

Question Order Effects on a General Chemistry Concept Inventory

Abstract:

During the development of a general chemistry concept survey, interviews demonstrated that students used distinctly different problem-solving strategies to answer two survey questions, one verbally-based and one pictorially-based, despite the fact that the questions were both designed to test the same concept of strong versus weak acids. Alternate versions of the concept survey were administered, with the order of the pictorial and verbal questions reversed. A significant ordering effect was observed in the questions of interest, and the incorrect answer choices that became better or worse distractors were identified. Current findings, future directions, and practical implications for instructors and researchers will be discussed.

Question-Order Effects in a Chemistry Concept Inventory

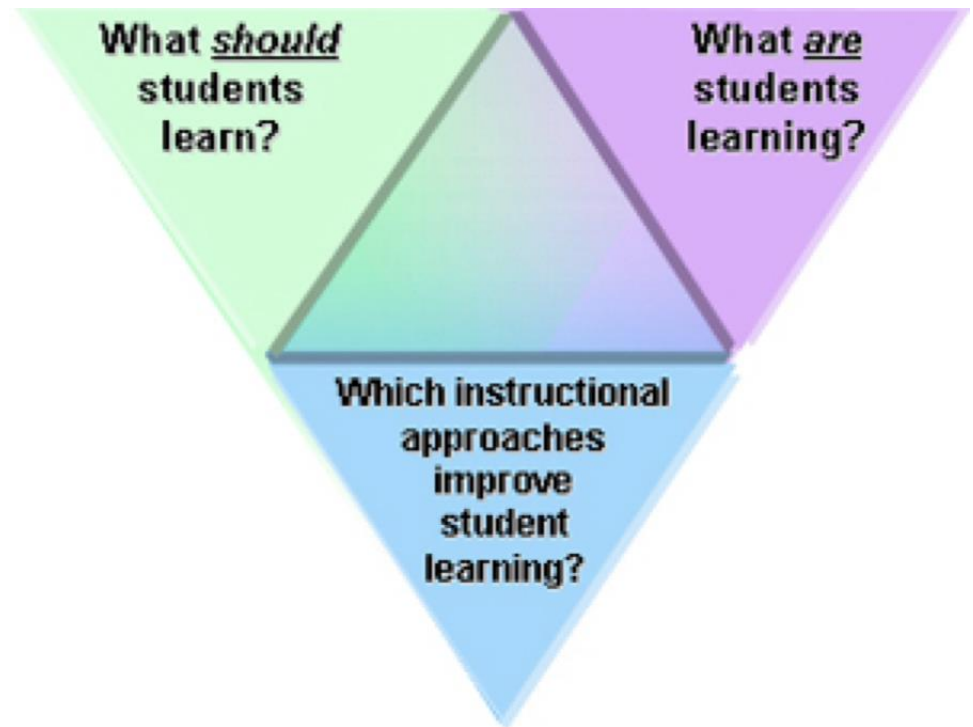


Travis Lund

Background: SEI

CU's **Science Education Initiative**: a university-supported, 5-year, \$5M project to improve how we teach science to all undergraduate students

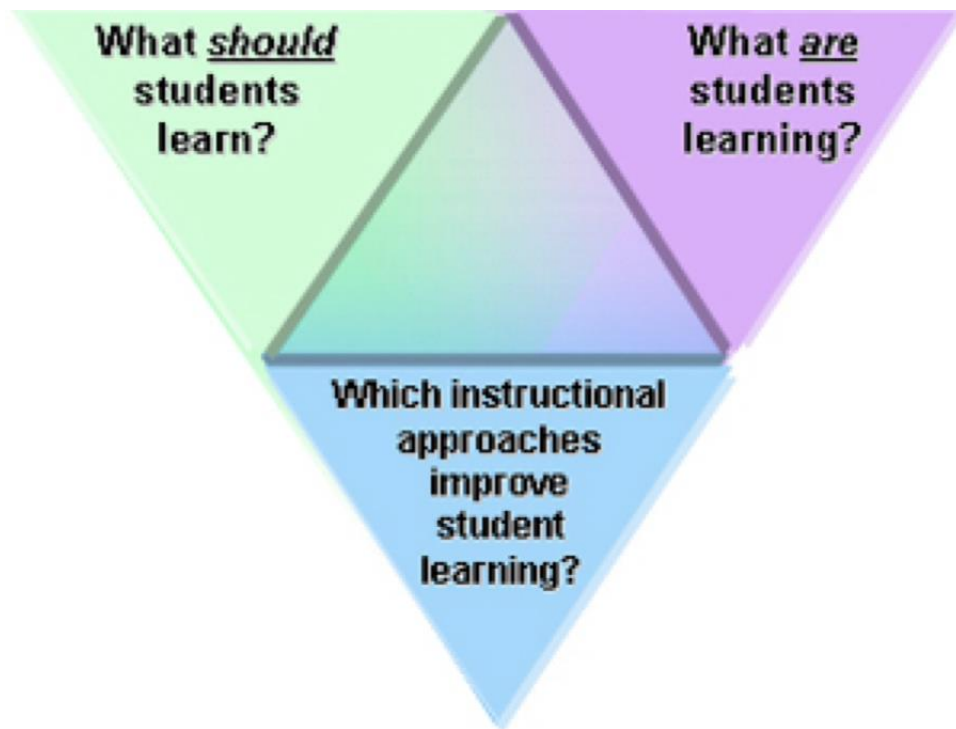
- Focus on achieving sustained, departmental-wide change
- Rely heavily on relevant science education research and technology
- Science Teaching Fellows (STFs): science education specialists with a Masters/PhD in specific departmental disciplines, supported by SEI to work with faculty and departments



Background: SEI in Chemistry

Dr. Laurie Langdon: STF for the Dept. of Chemistry & Biochemistry

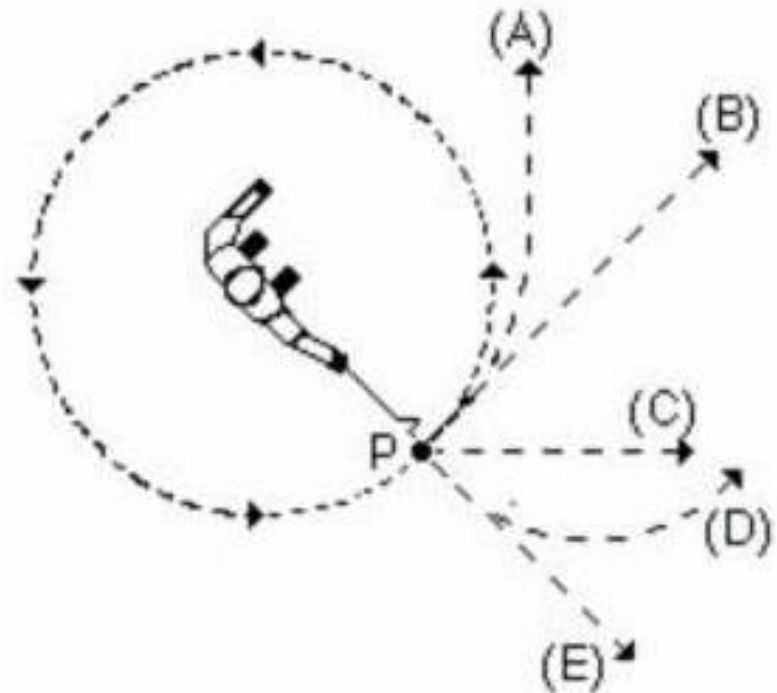
- Developed a **concept inventory** targeting CU's General Chemistry II learning goals
- Used as a pre-/post-test in general chemistry courses.
- 20-question multiple-choice test designed particularly to examine student understanding in the area of acid/base and solution chemistry.



Background: Concept Inventories

- Probes conceptual understanding, not problem-solving skills, etc
- Research-based questions and answers
 - Answer “distractors” are often common misconceptions
 - Validity and reliability testing
- Item difficulty values of 30-70%

- 1985: Force Concept Inventory, in physics
- Today: chemistry, biology, geoscience, astronomy, statistics, etc



Introduction

- During the concept inventory development, qualitative student interviews were performed with a student working through the test aloud for the interviewer
- Students used distinctly different problem-solving strategies or “mental frameworks” to answer certain questions on the test, although the questions were all designed to test the same concept (the relationship between pH, concentration, and acid/base strength).
 - Questions 9 and 10 are **pictorial**: P_a and P_b
 - Question 18 is **verbal**: V
- Is there an “ordering effect” in these questions?

Questions 9 and 10 refer to the drawings below. Four beakers each contain an acid. Molecular-level drawings in the circles to the right represent the particles in solution. Water molecules are not shown. Each beaker contains the same volume of solution.

KEY:



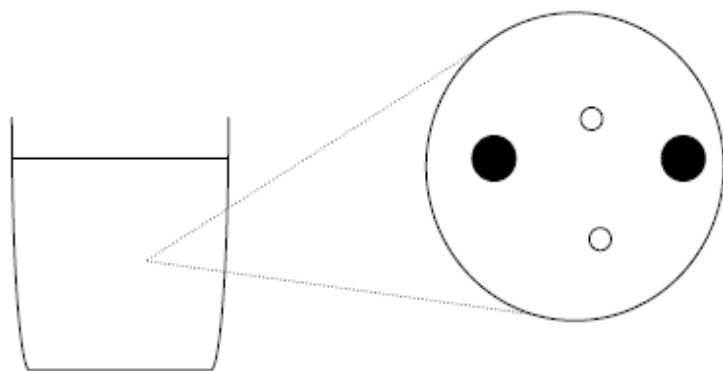
= HA (undissociated acid)



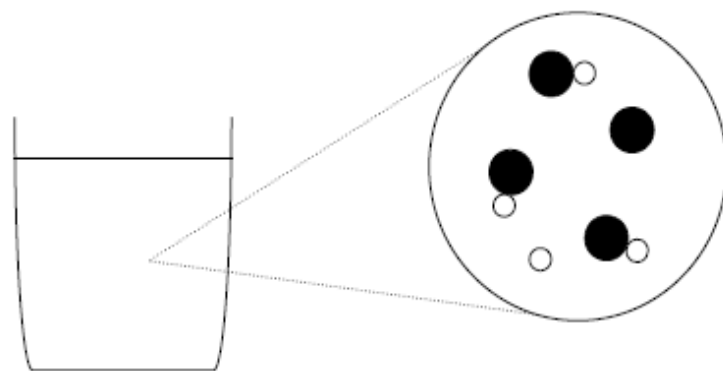
= A⁻



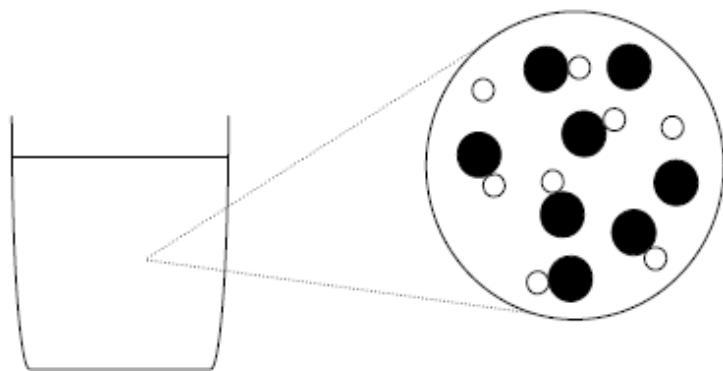
= H⁺ (or H₃O⁺)



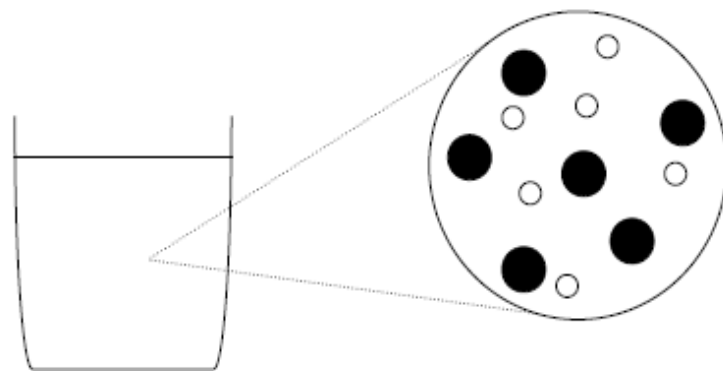
Beaker **A**



Beaker **B**



Beaker **C**



Beaker **D**

KEY:



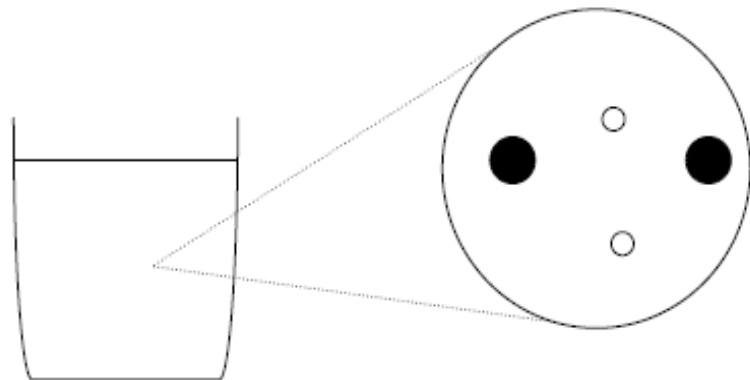
= HA (undissociated acid)



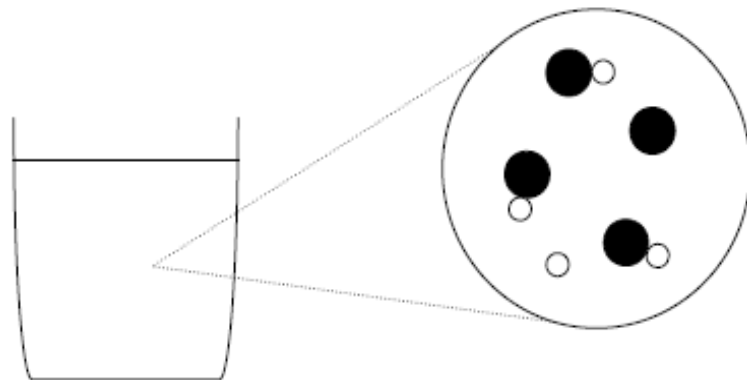
= A⁻



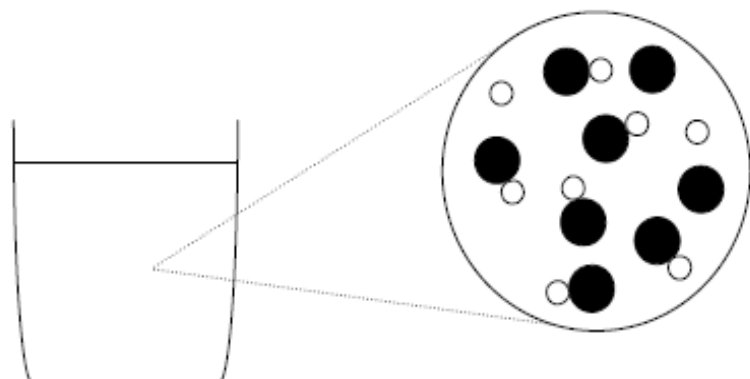
= H⁺ (or H₃O⁺)



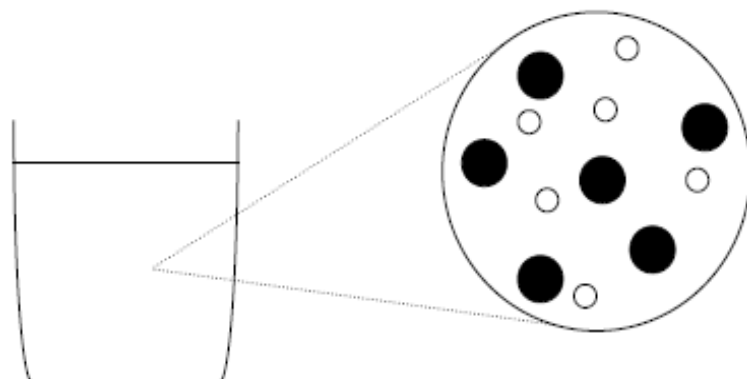
Beaker **A**



Beaker **B**



Beaker **C**



Beaker **D**

9. Which statement is **true**?

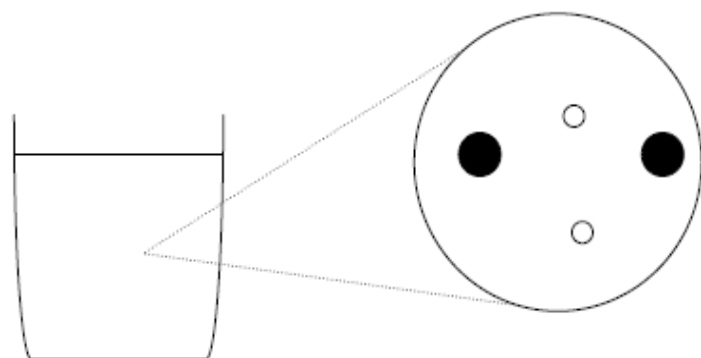
- a. Beakers **A** and **B** contain the weakest acids.
- b. The acid contained in Beaker **C** could be HCl (aq), which is a strong acid.
- c. The most concentrated acid is contained in Beaker **D**.
- d. The acid contained in Beaker **A** could be HNO₃ (aq), which is a strong acid.

KEY:

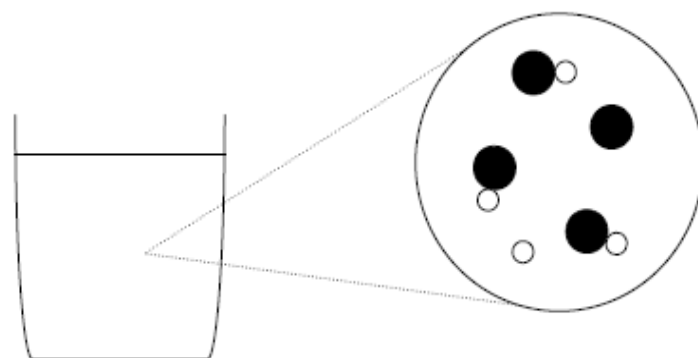
●○ = HA (undissociated acid)

● = A⁻

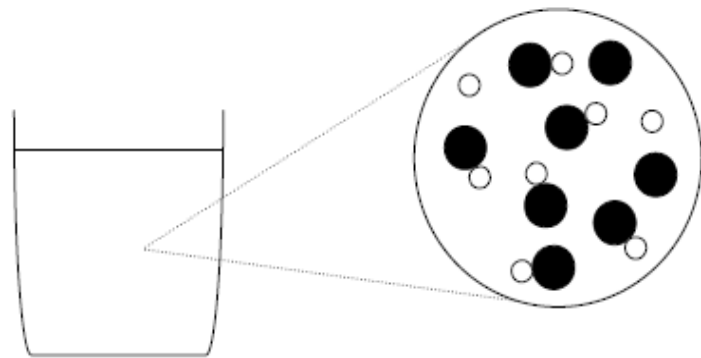
○ = H⁺ (or H₃O⁺)



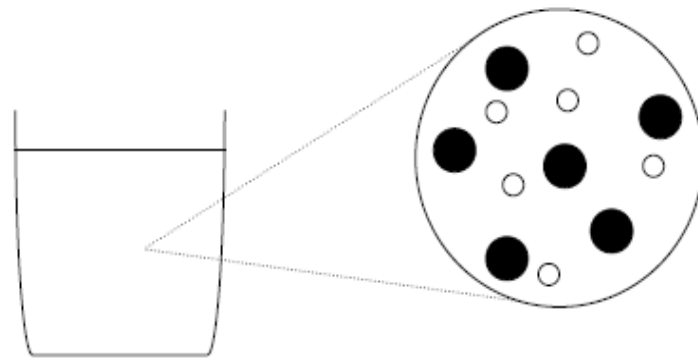
Beaker **A**



Beaker **B**



Beaker **C**



Beaker **D**

10. You measure the pH in each beaker. Order the beakers from lowest to highest pH.

Lowest pH → Highest pH

- a. Beaker **D** < Beaker **A** = Beaker **C** < Beaker **B**
- b. Beaker **A** < Beaker **B** < Beaker **D** < Beaker **C**
- c. Beaker **B** < Beaker **A** = Beaker **C** < Beaker **D**
- d. Beaker **D** < Beaker **A** < Beaker **C** < Beaker **B**

9. Which statement is **true**?

- a. Beakers **A** and **B** contain the weakest acids.
- b. The acid contained in Beaker **C** could be HCl (aq) , which is a strong acid.
- c. The most concentrated acid is contained in Beaker **D**.
- d. The acid contained in Beaker **A** could be $\text{HNO}_3 \text{ (aq)}$, which is a strong acid.

10. You measure the pH in each beaker. Order the beakers from lowest to highest pH.

- Lowest pH \rightarrow Highest pH
- a. Beaker **D** < Beaker **A** = Beaker **C** < Beaker **B**
 - b. Beaker **A** < Beaker **B** < Beaker **D** < Beaker **C**
 - c. Beaker **B** < Beaker **A** = Beaker **C** < Beaker **D**
 - d. Beaker **D** < Beaker **A** < Beaker **C** < Beaker **B**

18. What always distinguishes a weak acid from a strong acid?

- a. A weak acid doesn't dissociate much in water; strong acids completely dissociate.
- b. A weak acid is more dilute than a strong acid.
- c. A weak acid has a higher pH than a strong acid.
- d. Statements **a** and **c** are both characteristics that distinguish weak acids from strong acids.
- e. Statements **a**, **b**, and **c** are all characteristics that distinguish weak acids from strong acids.

Question-Order Effects

- Variation in question responses due solely to question context or placement

1. First Question

2. Second Question

3. Third Question

Question-Order Effects

- Variation in question responses due solely to question context or placement

Priming

1. First Question

2. Second Question

3. Third Question

1. First Question

2. Second Question

3. Additional Question

4. Third Question

Question-Order Effects

- Variation in question responses due solely to question context or placement

Location

1. First Question

2. Second Question

3. Third Question

1. First Question

2. Second Question

.

.

.

.

.

18. Third Question

Question-Order Effects

- Variation in question responses due solely to question context or placement

Transfer

1. First Question

2. Second Question

3. Third Question

1. Third Question

2. Second Question

3. First Question

Question-Order Effects

- Variation in question responses due solely to question context/placement
- Precedent for question-order effects can be found extensively in surveys and political polling, and recently in physics concept inventories.

More Overall Dissatisfaction When Asked After Bush Approval

Asked first	<i>Overall satisfaction</i>	<i>%</i>	<i>Bush approval</i>	<i>%</i>
	Satisfied	17	Approve	25
	Dissatisfied	78	Disapprove	67
	Don't know	<u>5</u>	Don't know	<u>8</u>
		100		100
Asked second	<i>Bush approval</i>		<i>Overall satisfaction</i>	
	Approve	24	Satisfied	9
	Disapprove	68	Dissatisfied	88
	Don't know	<u>8</u>	Don't know	<u>3</u>
		100		100
N		766		723

PEW RESEARCH CENTER Dec. 2008.

Question-Order Effects

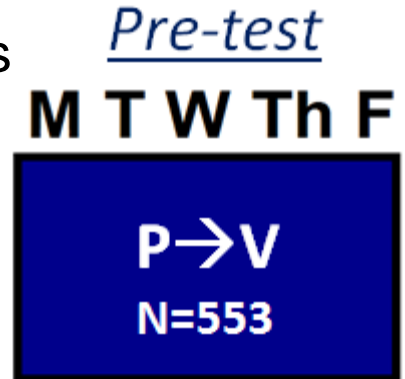
- Variation in question responses due solely to question context/placement
- Precedent for question-order effects can be found extensively in surveys and political polling, and recently in physics concept inventories.
- Educators often assume (implicitly or explicitly) that students' scores on a given question will not be influenced by context:
 - Constructing exams/inventories by drawing from a bank of questions
 - Randomizing the question order to produce alternate exam/inventory versions (to combat cheating)
 - Truncating, editing, or combining, existing assessments
 - Comparing scores on questions administered in different contexts

Question-Order Effects

- Variation in question responses due solely to question context/placement
- Precedent for question-order effects can be found extensively in surveys and political polling, and recently in physics concept inventories.
- Educators often assume (implicitly or explicitly) that students' scores on a given question will not be influenced by context:
 - Constructing exams/inventories by drawing from a bank of questions
 - Randomizing the question order to produce alternate exam/inventory versions (to combat cheating)
 - Truncating, editing, or combining, existing assessments
 - Comparing scores on questions administered in different contexts
- It is important to establish **whether, and to what extent, question-order effects may impact concept inventories** in chemistry and other disciplines.

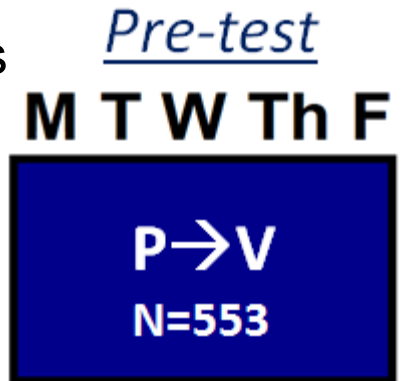
Methods

- 2nd semester general chemistry course; N=650 students
- **Pre-test** administered during first week of the semester, before the relevant material had been covered in class; pictorial questions appeared early, and the verbal question appeared late (“**P→V**”).



Methods

- 2nd semester general chemistry course; N=650 students
- **Pre-test** administered during first week of the semester, before the relevant material had been covered in class; pictorial questions appeared early, and the verbal question appeared late (“**P→V**”).



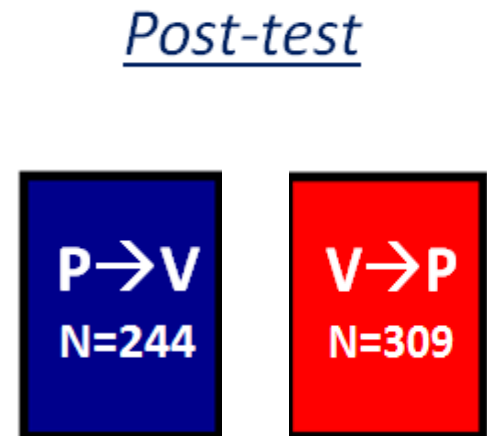
- **Post-test** administered near the end of the semester, long after the relevant material had been covered:

 **Post_{P→V} – Control** version maintained the original pre-test order:

- P (pictorial question) appeared early
- V (verbal question) appeared late

 **Post_{P→V} – Treatment** version reversed the positions of these questions:

- V appeared early
- P appeared late



Methods

The research question is one of causal estimation:

What is the effect of the *treatment* (reversing the order of P and V on the test) on the student responses to questions P and V?

- **Post-test** administered near the end of the semester, long after the relevant material had been covered:

Post-test

 **Post_{P→V} – Control** version maintained the original pre-test order:

- P (pictorial question) appeared early
- V (verbal question) appeared late

 **Post_{P→V} – Treatment** version reversed the positions of these questions:

- V appeared early
- P appeared late

P→V
N=244

V→P
N=309

Methods

- For logistical reasons, concept inventory post-test versions were not assigned to students randomly.

Pre-test

M T W Th F

$P \rightarrow V$ N=553				
----------------------------	--	--	--	--

Post-test

M T W Th F

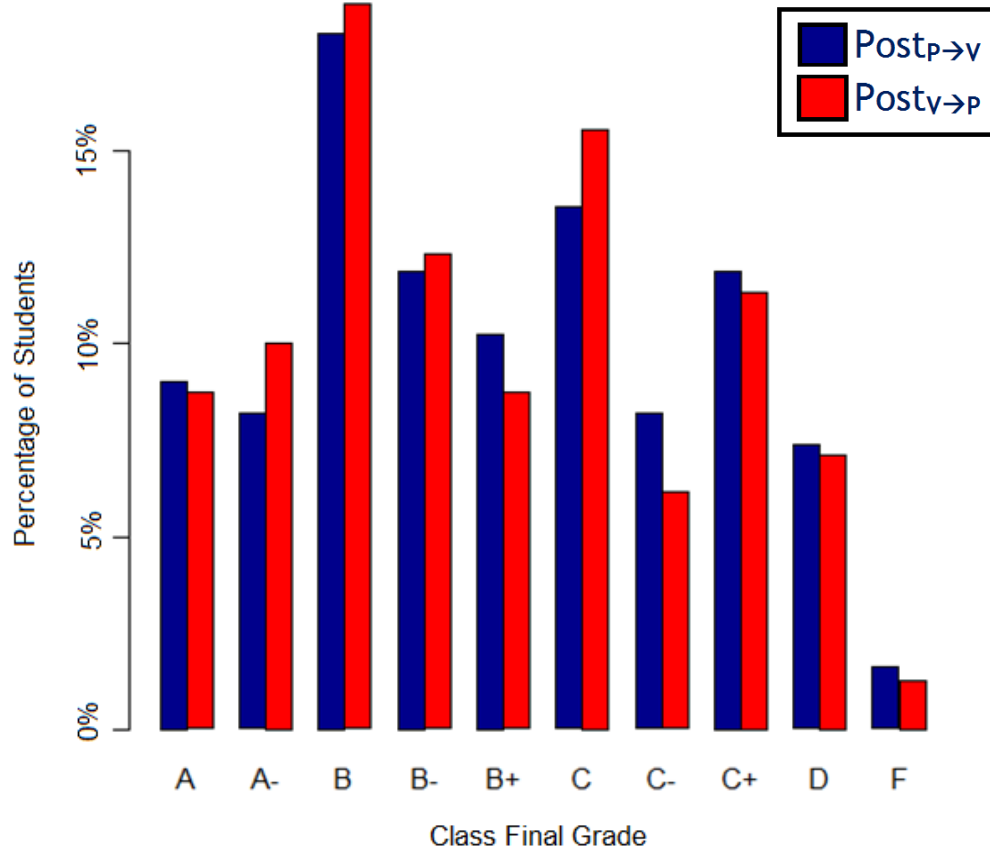
$P \rightarrow V$ N=244	$V \rightarrow P$ N=309
----------------------------	----------------------------

 $\text{Post}_{P \rightarrow V}$ – Administered early in the week

 $\text{Post}_{P \rightarrow V}$ – Administered late in the week

- Establishing that these two sets of students are comparable will be important for establishing the **internal validity** of any observed ordering effect on student performance.

Controls



	Post _{P→V}	Post _{V→P}
Mean GPA	2.62	2.85
SD	0.89	0.87
N	244	309

Figure 1. Grade Distributions by Post-test Version. Final class grades are comparable between Post_{P→V} and Post_{V→P}.

Controls

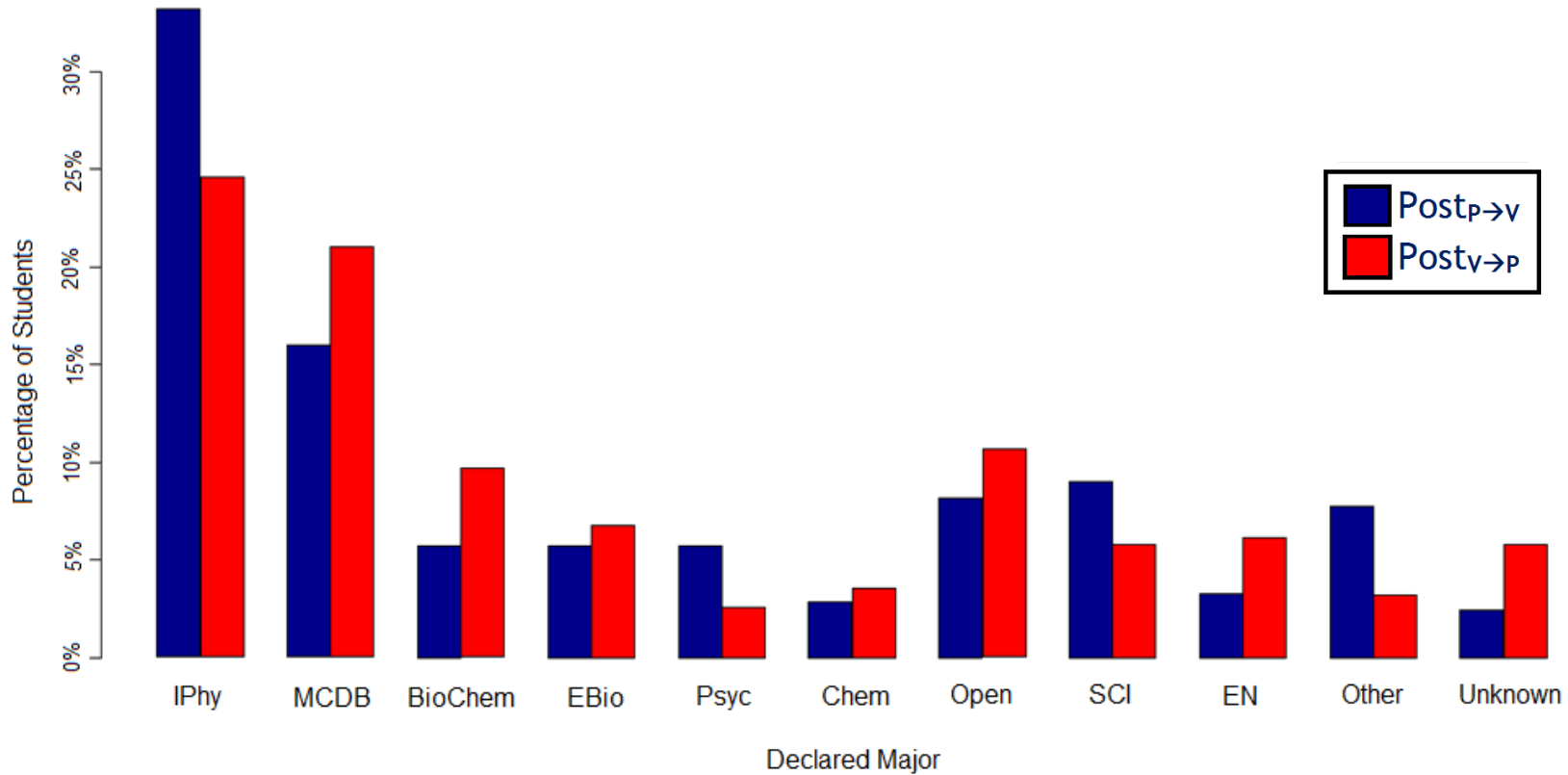


Figure 2. Declared Major by Post-Test Version. Approximately 70% of students declared their major as IPhy (Integrated Physiology), MCDB (Molecular, Cell, and Developmental Biology), BioChem (Biochemistry), EBio (Ecology and Evolutionary Biology), Psyc (Psychology), or Chem (Chemistry). The remaining 30% of students were categorized as Open (no declared major), SCI (other physical sciences, including physics, geology, environmental science, astronomy, and others), EN (all engineering majors), OTHER (art, business, and humanities majors), and UNKNOWN (major was not known).

The largest differences are for students declaring IPhy (8%), MCDB (5%), Other (5%), and Biochem (4%); all other differences are $\leq 3\%$.

Controls

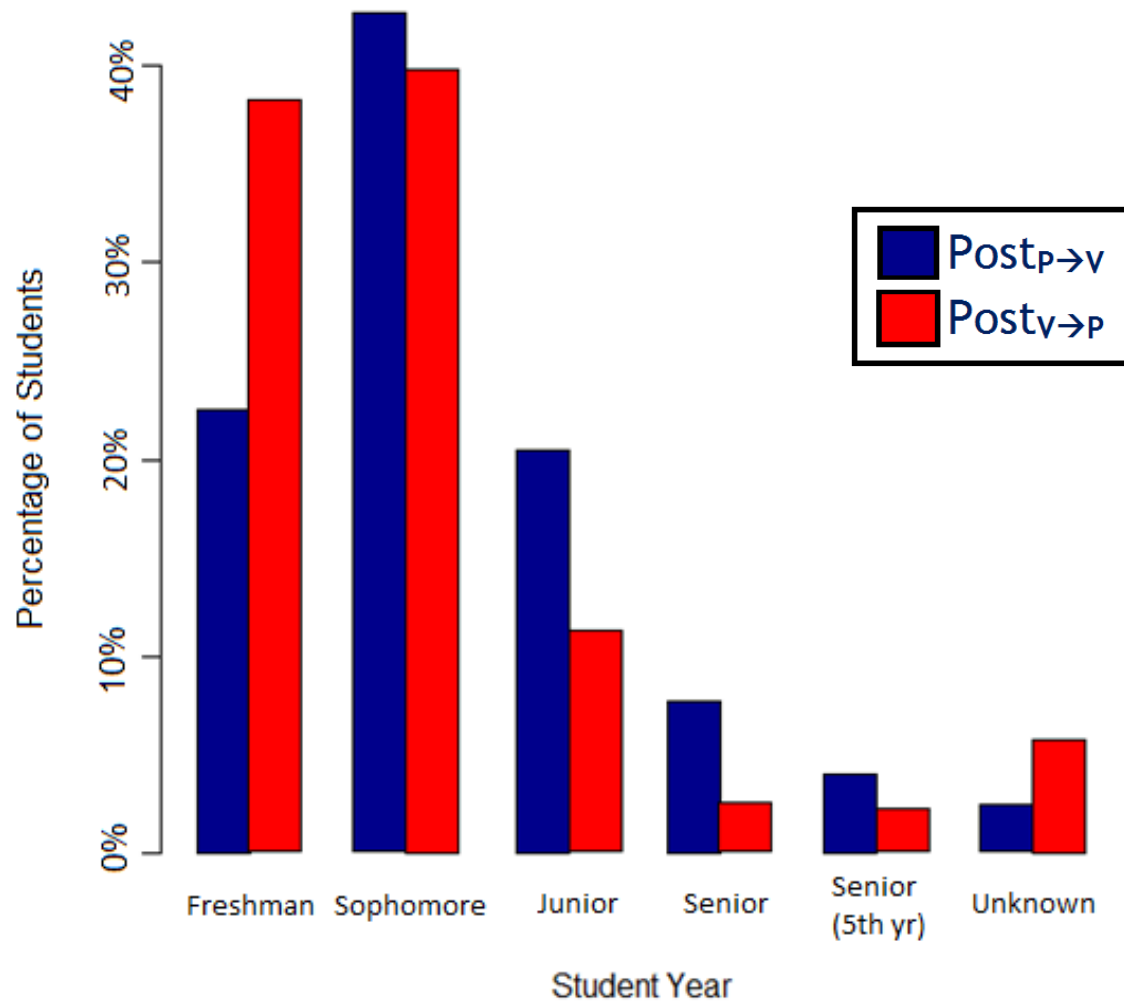


Figure 3. Year in College by Post-test Version. The largest differences are between freshmen (15%) and juniors (9%); all other differences are $\leq 5\%$.

Controls

Table 1. Concept Inventory Scores by Post-test Version

	Post_{P→V}		Post_{V→P}		<i>Difference</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
Pre-test	35%	14%	39%	14%	4%
Post-test, Q1-8	64%	21%	65%	21%	1%

Results

1) We observed an ordering effect of 8-9% upon reversing the order of the pictorial and verbal questions.

Question Accuracy by Post-test Version

	<u>Post_{P→V}</u>	<u>Post_{V→P}</u>	Difference
P _A	48%	56%	+8%
P _B	27%	29%	+2%
V	39%	30%	-9%

Results

1) We observed an ordering effect of 8-9% upon reversing the order of the pictorial and verbal questions.

Question Accuracy by Post-test Version

	<u>Post_{P→V}</u>	<u>Post_{V→P}</u>	Difference
P _A	48%	56%	+8%
P _B	27%	29%	+2%
V	39%	30%	-9%

- Low effect in P_b:
 - Context
 - Difficulty
 - Question “nature”

Results

1) We observed an ordering effect of 8-9% upon reversing the order of the pictorial and verbal questions.



Question Accuracy by Post-test Version

	<u>Post_{P→V}</u>	<u>Post_{V→P}</u>	Difference
P _A	48%	56%	+8%
P _B	27%	29%	+2%
V	39%	30%	-9%

Is this difference significant?

Results

Contextual Question Accuracy by Post-test Version

	Q4	Q5	Q6	Q7	Q8	P _A	P _B	Q11	Q12	Q13	Q14
 <u>Post_{p→v}</u>	64	84	72	59	58	48	27	64	54	55	38
 <u>Post_{v→p}</u>	64	83	75	59	60	56	29	68	56	56	38
<i>Effect</i>	0	-1	+3	0	+2	+8	+2	+4	+2	+1	0

Note: All numbers are percentages. Questions are numbered as in Post_{p→v}.

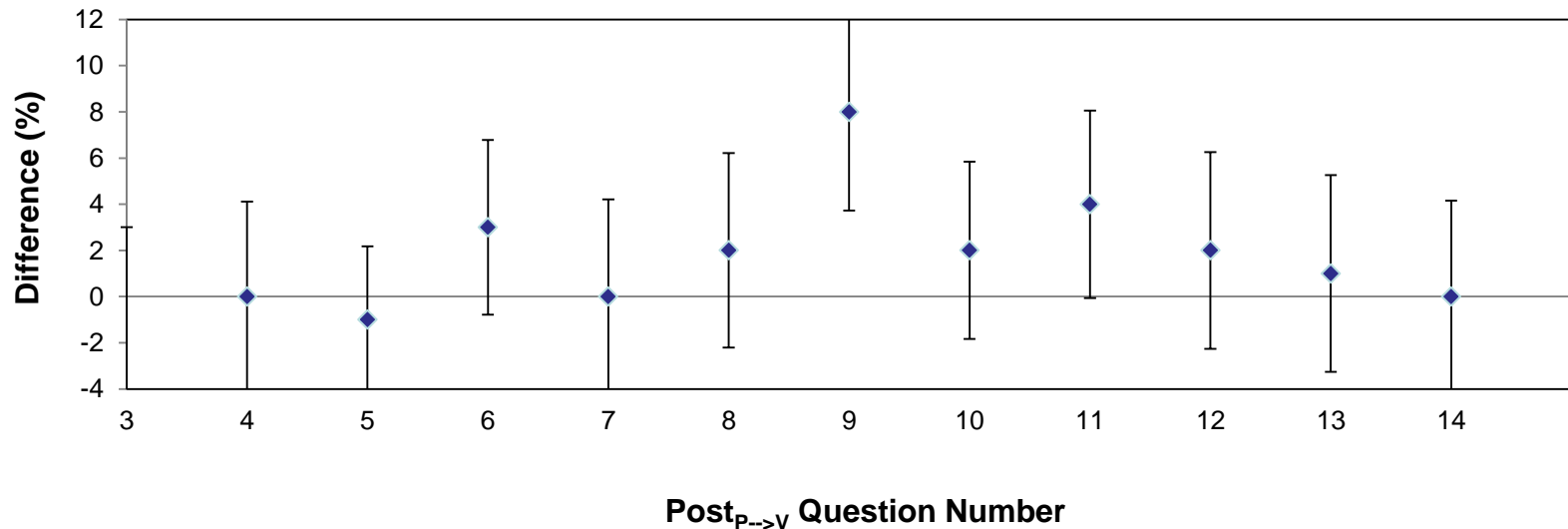
Results

Contextual Question Accuracy by Post-test Version

	Q4	Q5	Q6	Q7	Q8	P _A	P _B	Q11	Q12	Q13	Q14
■ <u>Post_{p→v}</u>	64	84	72	59	58	48	27	64	54	55	38
■ <u>Post_{v→p}</u>	64	83	75	59	60	56	29	68	56	56	38
Effect	0	-1	+3	0	+2	+8	+2	+4	+2	+1	0

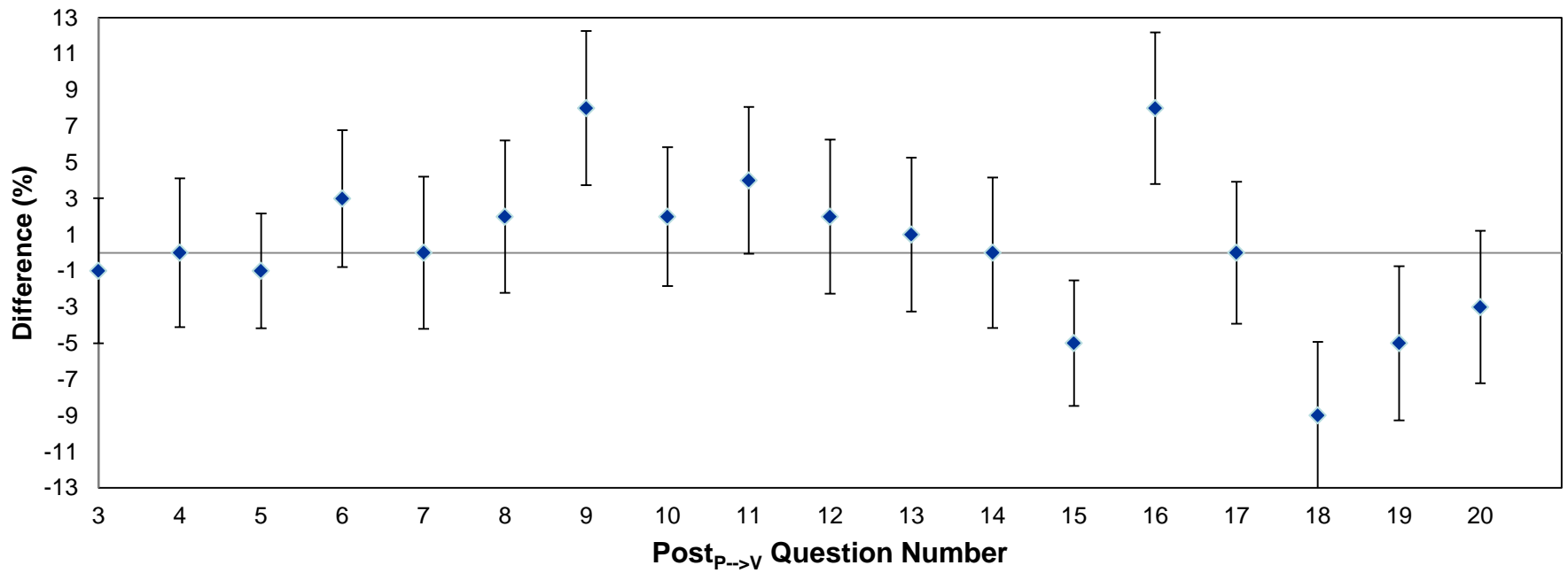
Note: All numbers are percentages. Questions are numbered as in Post_{p→v}.

Differences in Question Accuracy by Post-test Version



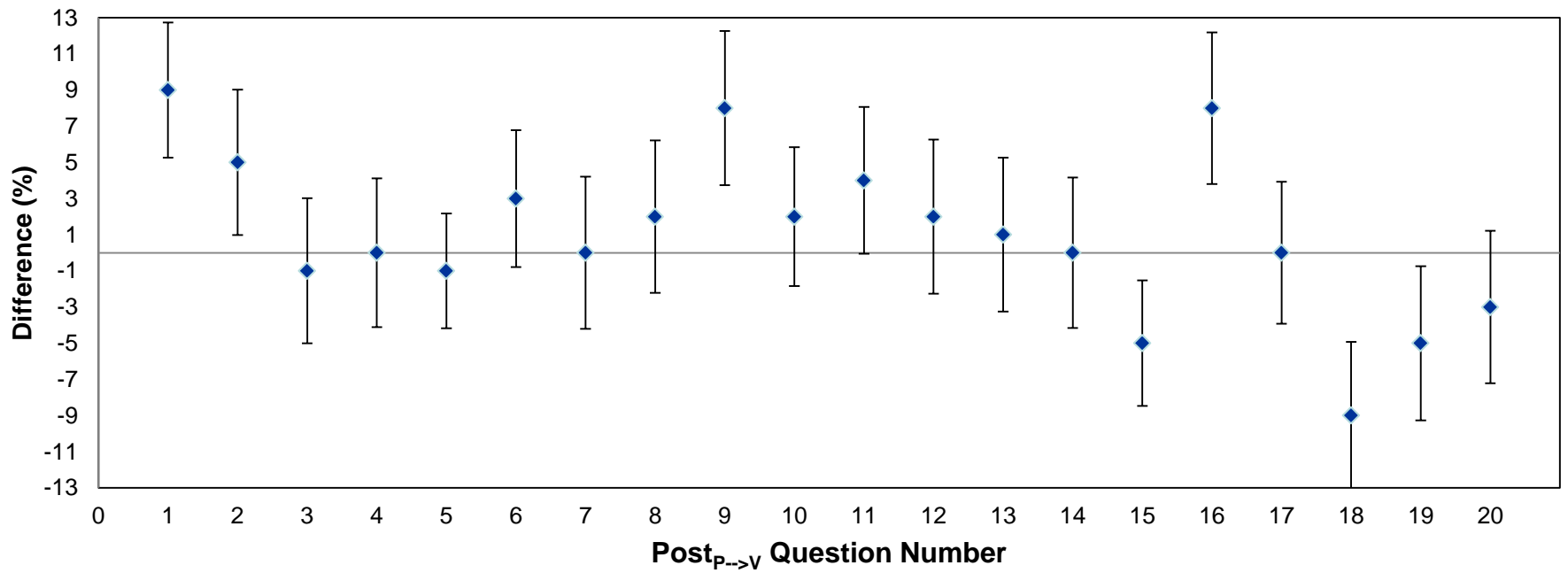
Results

Differences in Question Accuracy by Post-test Version



Results

Differences in Question Accuracy by Post-test Version



Results

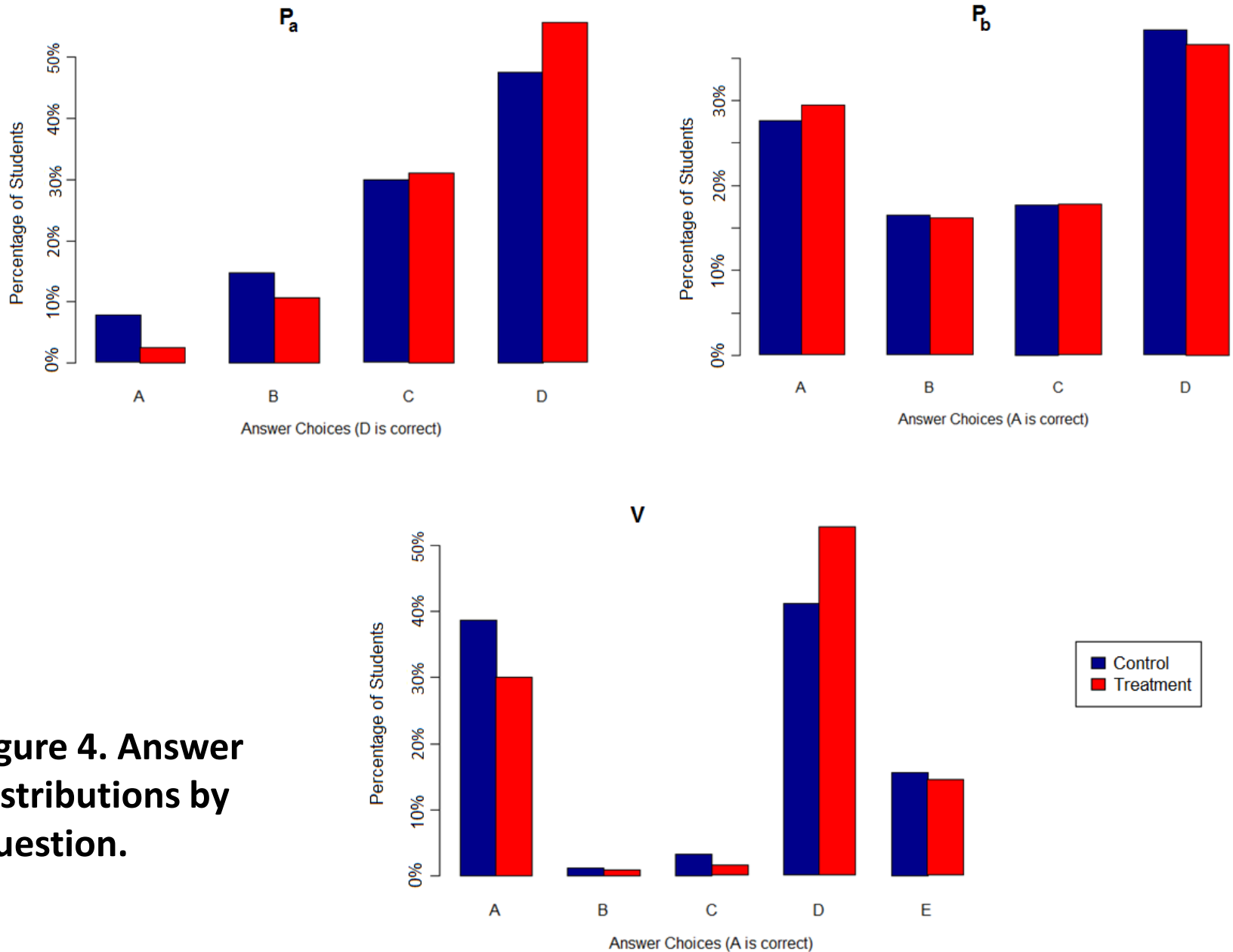


Figure 4. Answer Distributions by Question.

Conclusions

- 1) **We observed a statistically significant ordering effect of 8-9% upon reversing the order of the pictorial and verbal questions.**
- 2) **We identified certain answer choices that became better or worse distractors following the question reordering.**

Implications

- Educators and researches should not assume that student scores on any specific question are independent of context. Caution should be used when:
 - Truncating, editing, or combining existing assessments.
 - Constructing assessments by drawing from a bank of questions.
 - Randomizing question order to produce alternate assessment versions (e.g., to combat cheating).
 - Comparing student performance on questions, even if identically-worded, if those questions were presented in different contexts.

Future Work

- Is the observed ordering effect due to “priming” by the prior question (or questions), or due to question “location” in the test (early vs. late)?
- Are students “transferring” knowledge from the earlier question to the later?
- What role do question difficulty, question type (P vs. V), or other factors play in this effect?
- What conceptual factors underlie the differences in student approaches to the pictorial vs. verbal questions, or the systematic shifts in distractor selection?
- Are these factors inherent to the questions, or due to priming or other factors?



Acknowledgments



Dr. Laurie Langdon
Dr. Derek Briggs
Dr. Robert Kuchta
Dr. Marilyne Stains

Kuchta Lab
Stains Lab



Does the Answer Order Matter on Multiple-Choice Exams?

Joel Tellinghuisen* and Michelle M. Sulikowski

Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu

Journal of Chemical Education • Vol. 85 No. 4 April 2008

Does the Answer Order Matter on Multiple-Choice Exams?

Joel Tellinghuisen* and Michelle M. Sulikowski

Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu

Journal of Chemical Education • Vol. 85 No. 4 April 2008

1. First Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

2. Second Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

3. Third Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

Does the Answer Order Matter on Multiple-Choice Exams?

Joel Tellinghuisen* and Michelle M. Sulikowski

Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu

Journal of Chemical Education • Vol. 85 No. 4 April 2008

1. First Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

2. Second Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

3. Third Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

1. Third Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

2. Second Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

3. First Question

- A. Answer option A
- B. Answer option B
- C. Answer option C

Does the Answer Order Matter on Multiple-Choice Exams?

Joel Tellinghuisen* and Michelle M. Sulikowski

Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu

Journal of Chemical Education • Vol. 85 No. 4 April 2008

1. First Question
 - A. Answer option **A**
 - B. Answer option **B**
 - C. Answer option **C**
2. Second Question
 - A. Answer option **A**
 - B. Answer option **B**
 - C. Answer option **C**
3. Third Question
 - A. Answer option **A**
 - B. Answer option **B**
 - C. Answer option **C**

1. First Question
 - A. Answer option **C**
 - B. Answer option **B**
 - C. Answer option **A**
2. Second Question
 - A. Answer option **B**
 - B. Answer option **A**
 - C. Answer option **C**
3. Third Question
 - A. Answer option **A**
 - B. Answer option **C**
 - C. Answer option **B**

Does the Answer Order Matter on Multiple-Choice Exams?

Joel Tellinghuisen* and Michelle M. Sulikowski

Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu

To judge whether the differences are statistically significant, we assess their uncertainties using binomial statistics (1). Each trial is assumed to have only two outcomes, right or wrong, with blank answers included among the latter. If p represents the probability for a specified outcome and there are N trials, then the variance σ^2 is given by

$$\sigma^2 = Np(1 - p) \quad (1)$$

Since the mean is $\mu = Np$, simple error propagation gives for p the standard deviation σ/N . Expressing the results as percent right (or wrong), the uncertainty is thus

$$\sigma_{\%} = \left[\frac{(1 - p)p}{N} \right]^{\frac{1}{2}} \times 100\% \quad (2)$$

For the computation of $\sigma_{\%}$, we take p as the fraction wrong on a given question, for all papers of both test versions combined. We then plot the difference, $p(\text{blue}) - p(\text{gray})$, both expressed as percentages. The uncertainty in this difference is, from the usual rules for error propagation in sums (1), the square root of the sum of $(\sigma_{\%})^2$ for the two quantities.

Pyrimidine O², N-3 and O⁴/N⁴ play essential but asymmetric roles for efficient dNTP incorporation by Klenow fragment.

