Enhancing the Geoscience Pedagogical Landscape: An Integrative Strategic Approach to Heighten Student Learning

M.W. Cole (Melanie Will-Cole)
Fellow U.S Army Research Laboratory, Emerita
Central New Mexico Community College (CNM)
MSE-Dept. of Earth & Planetary Science
525 Buena Vista Dr SE, Albuquerque, NM 87106
TEL: (505) 433-2444
Email Melanie.willcole@gmail.com
Outline of Presentation

• Introduction

• Motivation and Background
  ✓ New pedagogical methods to enhance learning
  ✓ 2-stage exams
  ✓ Utilization and benefits

• Experimental Study 2-Stage Exam
  ✓ Design & implementation
  ✓ Tailorability

• Results & Discussion
  ✓ Analysis of the trends in student performance for individual-group exam
  ✓ Analysis of student learning from group exam
  ✓ Evaluation of student survey assay

  ▪ Summary & Conclusions

  ▪ Next Steps & Future Work
Current Trend in Higher Education

• Employ active learning methods to enhance student learning
  (used in classroom & often used as Formative Assessment)

• Research Based Methods of collaborative learning
  ✓ Peer instruction
  ✓ Think-Pair-Share
  ✓ Group problem solving
  ✓ Interactive lectures
  ✓ Engineered laboratory simulations

• Common Feature include
  ✓ Intense student engagement
  ✓ Instructor guidance
  ✓ Collaborative learning
    (students develop thinking & learning via peer interaction-immediate feedback)
Identify New Pedagogical Methods to Enhance Student Learning

- Active learning STEM pedagogies are highly researched (publications) & becoming widely distributed (used)
- Most science classes incorporate testing (summative assessment)
- Exams as tool to gauge student achievement of the course learning objectives
  - Learning is not emphasized during exam
  - Takes place as a by-product after the exams
- Test methodologies still remain quite traditional
- Traditional testing format
  - Summative assessment of individual performance
  - Students works alone → solves problems in isolation
  - Intense engagement with material with little contribution to learning (acquiring new ideas)
  - Receives delayed feedback
  - Does not support collaborative learning
  - Often promotes anxiety → false positive/negatives for performance
What is a 2-Stage Exam?

- Goal is to turn test sessions into learning experiences
- Students take same test twice during class period but in different settings
  - 1st individual Exam
  - 2nd take exam collaboratively in pairs or groups
- Scoring calculate based on a weighted average of the individual part (90-60%) and the group part (10 - 40%) of the exam
- If group score is < individual score final score is based solely on individual exam

“Collaborative Part of Exam Will Not Lower Students Grade”

Motivation & Background

- Active learning methods promotes enhanced student learning
- Why not use test methods which foster learning
  
  Learning can take place during exams vs. recall of material or concepts
- Collaborative Testing (team based learning)
  
  Two Stage Exams/Group Test/Pyramid Exam
- Valuable for learning & synergistic with in-class collaborative pedagogies
**Motivation & Background: Characteristics of Exam Formats**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Traditional Exam Format</th>
<th>Two-Stage Exam format</th>
</tr>
</thead>
</table>
| Nature of Exam        | • Individual focused  
                      • Summative assessment of performance (individual)                                      | • Individual + Group collaborative  
                      • Summative assessment of performance (individual)                                      |
| Students              | • Passive                                                                | • Active-intense engagement by students                                                 |
| Problem Solving       | • Isolation                                                              | • Collaborative discussions                                                            |
| Ambience              | • Quiet                                                                  | • Noisy                                                                                 |
| Learning Environment  | • Competitive                                                            | • Supportive, collaborative                                                              |
| Learning Opportunities| • Limited  
                      • Intense engagement with subject matter                                  | • Multiple  
                      • Intense engagement in student learning, Share/discuss ideas/generate new ideas |
| Students Role         | • Student (learner only)                                                   | • Learner (student) & Teacher (explain to others) – Students serve as facilitator       |
| Feedback              | • Delayed                                                                | Immediate through interactions with peers                                               |
| Results               | • Individual grade attained  
                      • Is learning really assessed???                                                     | • Improved learning (self + group-acquisition of knowledge by students)                        |
| Consequences & Benefits|                                                                                             | • Increased retention of subj. matter content                                          |
|                       |                                                                                             | • Decreased drop out rates                                                             |
|                       |                                                                                             | • Higher course enjoyment                                                              |
|                       |                                                                                             | • Increased collaborative skills                                                       |
|                       |                                                                                             | • Reduced test anxiety                                                                 |
|                       |                                                                                             | • Improved perception of course                                                        |
Motivation & Background

- Two-Stage Exams are not a new idea; but utilization is still relatively rare
- Few published studies measured the benefits on learning in science
  - Enhanced student learning
  - Decreased drop out rates
  - Increased attendance
  - Higher enjoyment if the course
  - Increased collaborative skills

- Leaders in the field University of British Columbia, Vancouver, BC, Canada, Carl Wieman
- Used in physics, nursing, medical-related courses. biology courses
Experimental Study & Methodology

- Central New Mexico Community College, 2-year college; mission is to educate students & prepare students for transfer to four-year institutions.

- Student population studied (Introductory Physical Geology Course)
  - Freshman-sophomore level
  - Mostly local NM based (few/none foreign nationalist students)
  - Mix of mature & recent HS grads, life experiences, most working ~part time
  - Mix of geoscience majors and non-majors (need a lab science course)

- Class attendance is mandatory (students allowed to miss only 15% of classes)
- Administered 2-stage exam to two sections of intro. physical geology course
- Class period of 75 minutes (used time allocation 1:1)
- Total number of students: Test #3 (n=42), Test #4 (n=40), Retest (n=35)

- The 1st part of the semester all the students took 2 individual in-class exams
- Exams #3 & 4 were administered as 2-stage exams;
  - Exams problem solving & concept demonstration/evaluation based exams
  - Test #3: relative-absolute time, earth history
  - Test #4: structural geology, seismology, oceanography
Exam Methodology
Tailorable

1. First stage of Exam ➔ Students takes exam solo
   • 25 questions, 30 minutes

   - Collect Individual exams
   - Students move into randomly pre-assigned groups (5 min)

2. Second stage of Exam
   • Groups of 3 - Randomly selected
   • 25 questions, 30 minutes
   • 1 exam per group
   • 3 Student observers monitor effectiveness of exam

   - Collect group exams
   - Review exam interactively with participation from the student groups
   - Student survey

Exam #3: Relative - Absolute Time & Earth History
Exam #4: Structural Geology, Seismology, Oceanography

Grading Methodology ➔ Tailorable
• Score Final = Score \(_{\text{Individual (I)}}\) + Score \(_{\text{Group (II)}}\)
• Scores are proportionally weighted to suit specific class/student situations
• Ensures students take both components seriously

3. Learning Exam ➔ Students takes same exam solo
   • 25 questions, 30 minutes
   • Administer 4-8 weeks after 2-stage test (this study 6 weeks)

Same Class Period

4 to 6 weeks later
Experimental Study & Methodology

Important Tailorable Considerations
1. Group selection method & number of students per group
2. Timing of each exam component
3. Grading → weight (Indiv.: Group) 90:10, 80:20, 70:30 etc.
4. Types and number of questions on exam etc.
5. Timing of learning test

Students Taking Collaborative Exam
Results & Discussion: Comparison of Individual & Group Scores

**Plot T3**

Exam Score (%) vs Students for Test #3.
- O Individual Exam
- ■ Group Exam

**Observation**
Group Score are higher

**Plot T4**

Exam Score (%) vs Students for Test #4.
- O Individual Exam
- ■ Group Exam

**Data**
PAIRED t-test_ test #3 WS + Montoya_ ind vs group test_7-2-2017
A-indiv Score
B-gp Score

PAIRED t-test_ test #4 WS + Montoya data ind-gp-retest_for t-test -mod7-1
A-indiv Test #4
B-gp test #4

**Data**
I-Student
Results & Discussion: Comparison of Individual & Group Scores

**Observation**
Group Score is higher than best Individual score.
Results & Discussion: Point Gain Individual - Group Scores

Plot-BAR_T3 Std vs pt gain & ind score

Test #3

Point Gain
-20 0 20 40 60 80 100

Students

Group Exam Score

Point Gain

Individual Exam

(Stacked Column Plot)

Group Exam ➔ Improved Scores

Plot-BAR_T4 Std vs pt gain & ind score 7-2-17

Test #4

Point Gain
-20 0 20 40 60 80 100

Students
Results & Discussion: Individual and Group Exam Scores

- Students plotted above ref. slope → showed improvement from Ind. to group exam
- Students with lower scores on Individual Exam gained more points in Group Exam
- Students with higher scores on Individual Exam gained less points in Group Exam

Y = 0.28642x + 63.48; R = 0.41595

Y = 0.70529x + 32.204; R = 0.68039
Results & Discussion: Paired t-test - Individual & Group Test Scores

**Paired t-Test -**
- **✓** Compare means on the same subjects over time or under different conditions
- **✓** Used to determine if two data sets are significantly different from each other

### Student t-test
- Is there a statistical difference between individual & group scores (the means)?
- Did the group exam make a difference?  t-test said yes!
- t-probability <0.0001
- t-test results demonstrated that the results are statistically different
- There is a statistically significant difference (improvement) in student performance between individual and the group test

<table>
<thead>
<tr>
<th>Values</th>
<th>Test #3 (n=42)</th>
<th>Test #4 (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>Maximum</td>
<td>84</td>
<td>94</td>
</tr>
<tr>
<td>Mean</td>
<td>56</td>
<td>56.85</td>
</tr>
<tr>
<td>Median</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>RMS</td>
<td>58.018</td>
<td>59.233</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.3528</td>
<td>16.479</td>
</tr>
<tr>
<td>Variance</td>
<td>235.71</td>
<td>271.57</td>
</tr>
<tr>
<td>Snd. Error</td>
<td>2.369</td>
<td>2.6056</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.33792</td>
<td>0.73771</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-1.0392</td>
<td>-0.39843</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test #3 (n=42)</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Std. Err.</td>
</tr>
<tr>
<td>Mean Difference</td>
</tr>
<tr>
<td>T-value</td>
</tr>
<tr>
<td>t-Probability</td>
</tr>
<tr>
<td>Correlation</td>
</tr>
<tr>
<td>Corr. Probability</td>
</tr>
</tbody>
</table>
Analysis of the trends in student performance level with respect to scores on group test

Who Benefited the Most?

Influence of Student Performance Level on Group Exam Scores

- Individual Exam → baseline for binning students by performance level
  
  Students → High, Mid & Low performance (Individual Exam scores)

  ✓ Test #3: High ≥ 64%, Mid 63-48%, Low <48% (equal #s in each category)
  ✓ Test #4: High ≥ 60%, Mid 59-46%, Low <45% (equal #s in each category)
Students binned by performance level based on Individual Exam Scores

- Test #3: High $\geq 64\%$, Mid 63-48\%, Low $< 48\%$ (equal #s in each category)
- Test #4: High $\geq 60\%$, Mid 59-46\%, Low $< 45\%$ (equal #s in each category)

- All students benefited from group exam (Score Improvement)
Influence of Student Performance Level on Group Exam Scores

Students → High, Mid & Low performance (Individual Exam scores)

- Binned Students by performance level based on Individual Exam Scores
  - Test #3: High ≥ 64%, Mid 63-48%, Low <48% (equal # in each category)
  - Test #4: High ≥ 60%, Mid 59-46%, Low <45%

- All students benefited from group exam (Score Improvement)
- Mid and Low Performers benefited most
Student Benefit by Performance Level → Box Plot Test #3

- High performers lower benefit from group exam
  - Improved mean score 11% (indiv. to group exam)
  - Moderate degree of dispersion (spread) in data (moderate box width)
Student Benefit by Performance Level → Box Plot Test #3

- High performers lower benefit from group exam
  - Improved mean score by 11% (indiv. to group exam)
  - Moderate degree of dispersion (spread) in data (moderate box width)

- Low Performers largest benefit from group exam
  - Improved mean score by 89% (indiv. to group exam)
  - Lowest degree of dispersion within test type data (narrow box width)

- Mid Level performers close to that of low performers
  - Improved mean score by 62% (indiv. to group exam)
  - Moderate degree of dispersion within test type data (moderate boxes width)
High performers lower benefit from group exam

- Improved mean score 10.8% (indiv. to group exam)
- Moderate degree of dispersion (spread) in data (moderate box size)
High Performers lower benefit from group exam
- Improved mean score 10.8% (indiv. to group exam)
- Moderate degree of dispersion (spread) in data (moderate box size)

Mid & Low Performers → largest benefit from group exam
- Mid Performers improved mean score 42% (indiv. to group exam)
- Low Performers improved mean score 44.6%
- Both have larger dispersion with in group scores (Group Effect)
- Both have minimal dispersion in the individual scores
All students benefited from group exam

Next Questions to Ask (Answer)

1. Did the students really “Learn” as a result of group test?

2. Which students benefited most? (high, mid-level or low performers)

- Retest administered 6 weeks after the 2-stage exam was given
- Individual test baseline assessment of student knowledge
- Retest (Learning Test) serves as a measure of student learning
- Did the collaborative part of the test improve individual student learning? (Knowledge Banking Effect)
**Student Learning:** Point gain/loss between Retest & Individual Test

- Retest administered 6 weeks after the 2-scage exam was given
- Retest serves as a measure of student learning
- All Students improved their scores on retest (post 2-stage exam) vs. Individual Test
Student Learning: Individual Exam vs. Retest Exam Scores

- Retest serves as a measure of student learning
- Students plotted above ref. slope → showed improvement from Ind. to retest exam
- Same slope, different y-intercept → most learning test scores moved upward
- Students improved their scores on retest (after 2-stage exam) vs. Individual Test

Y = 0.94419x + 10.309; R = 0.94018

Test #4
n=35
**Paired t-Test - Individual & Retest Scores (test # 4)**

**Paired t-Test -**

- Compare means on the same subjects over time or under different conditions
- Used to determine if two data sets are significantly different from each other

<table>
<thead>
<tr>
<th>Test #4 (n=35)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Exam</td>
</tr>
<tr>
<td>Mean</td>
<td>55.71</td>
</tr>
<tr>
<td>Variance</td>
<td>252.975</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.9052</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>2.68847</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-7.2</td>
</tr>
<tr>
<td>t-value</td>
<td>-7.7256</td>
</tr>
<tr>
<td>t-Probability</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.94018</td>
</tr>
<tr>
<td>Corr. Probability</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

**t-test results**

**Student t-test**

- Is there a statistical difference between Individual & Retest scores (means)?
- Did the group exam make a difference? t-test said yes!
- t-probability <0.0001 (demonstrates that the results are statistically different)
- There is a statistically significant difference (improvement) in student performance between individual and the retest scores
- Improvement on leaning test indicates
  - During the collaborative test the students acquired new knowledge
  - Learning was sufficient enough for students to remember it ~6 weeks later
All students benefited from group exam

Next Questions to Ask (Answer)

1. Did the students really Learn as a result of group test?

2. Which students benefited (learned) most? (high, mid-level or low performers)

- Retest administered 6 weeks after the 2-stage exam was given
- Retest primary measure of student learning
- Understand the effect of 2-Stage Exam on Student Learning as a function of student performance level
- Students binned by performance lever, high, Mid and Low
Influence of Student Performance Level on Retest Exam Score

Students binned by High, Mid & Low performance from Individual Exam scores

- All students learned from group exam (score improvement on retest)
Influence of Student Performance Level on Retest Exam Score

Students binned by High, Mid & Low performance from Individual Exam scores

- All students learned from group exam (score Improvement on retest)
- Mid and Low Performers benefited slightly more than high performing students
Student Benefit by Performance (Test #4)

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv. Mean</th>
<th>Group Mean</th>
<th>R-Test Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>74.5</td>
<td>82.6</td>
<td>80.3</td>
</tr>
<tr>
<td>Median</td>
<td>74</td>
<td>78</td>
<td>83</td>
</tr>
</tbody>
</table>

- **High Performers** least difference between Individual and Retest exam scores
  - Moderate degree of dispersion (spread) in data
  - Mean score improved by 11% on Retest

- **Low & Mid Performers** largest difference between Individual & Retest exam scores
  - Low degree of dispersion in data
  - Mean score improved by ~16% (Mid) & 16.6% (Low)

- Methods to improve low and mid performers learning are critical in higher education
  - Benefit directly result of peer teaching
  - Benefit due to greater motivation to learn material (better attitude about learning)
  - Students better prepared - “desire to look good amongst peers”
Student Evaluations
Student Observer Comments

- Most students were being productive while working in their groups.
- When one student did not understand another would step in & try to explain the concept.
- ~90% of students were discussing, asking questions & analyzing the test questions. 10% either did not study or were just agreeing on the answer w/o discussing.
- Students seemed to be enjoying and learning from one another.
- The groups seemed to create a positive learning atmosphere.
- A few groups had a predominant figure that who conducting the discussion, most others were on a parallel level.
- Having peer evaluation while taking the exam is reinforcing both the material known and understood as well as pointing out mistakes originated because something was not entirely comprehended.
- Some cases the discussions were not long because all the students readily agreed to the same answer.
Student Participant Comments

- The group exam helped me learn from other students & made me question their answers while being able to discuss their reasoning.
- My partners all contributed to the discussion.
- Thought I knew this stuff pretty well but the group exam showed me that I needed to think a bit more on certain questions.
- This experience helped to get a better understanding of relative & absolute dating.
- As a student I found it easier to communicate with other students about what we are learning in class than in the regular class.
- I know I am not doing stellar on my tests but now I have been exposed to a big positive learning style & you have really shown me that.
- The group test made me and my group more relaxed, sort of relived my test anxiety, seemed easier to learn from this group environment.
- Never had a group exam before, would like to see it used more, It was fun
**Student Survey Questions**

<table>
<thead>
<tr>
<th>Question</th>
<th>Agree</th>
<th>Disagree</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Contributed to the development of my ability to critically analyze a problem related to the subject matter through group collaborative discussions.</td>
<td>88%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>2. Group discussion often/sometimes caused me to re-evaluate some of the exam answers I had selected on the individual exam which I took prior to the group exam.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Aided my learning of the subject matter via the group discussion &amp; information exchange amongst my group members.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Collaborative discussion between me and my group members, served to make the problem solving component of the subject matter more explicit to me.</td>
<td>82%</td>
<td>6%</td>
<td>12%</td>
</tr>
<tr>
<td>5. My group most always interacted in a collaborative manner arriving at the answers to the exam questions through group discussion and consensus.</td>
<td>94%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>6. Group part of exam helped me to formulate &amp; explain my ideas.</td>
<td>94%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>7. Answers to the questions were formulated through contributions from most all the group members.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. My group members were respectful of the ideas and contributions put forward by others within the group.</td>
<td>94%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>
## Generating Science Relationships

<table>
<thead>
<tr>
<th>Question</th>
<th>Agree</th>
<th>Disagree</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collaborative discussions amongst me and my group member’s encouraged me to think critically, share my thoughts &amp; construct explanations regarding the subject matter problem solving questions posed on the exam.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. There are more opportunities to generate scientific ideas &amp; inquiry with respect to the problem solving questions in the group part of the 2-stage exam than in other classes which use traditional individual written exams.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I find myself asking “what would happen if…” science-type questions more often than in other courses which use traditional individual written exams.</td>
<td>94%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>4. I was able to discuss &amp; modify my ideas about the subject matter because my group members had explained concepts to me that I did not fully understand prior to taking the exam.</td>
<td>88%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>5. As a result of the group discussions, I found that I modified some of my initial ideas (with respect to my answers on the individual exam) regarding science relationships pertaining to the topic of absolute and relative time/oceanography, structural geol., seismology.</td>
<td>94%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>
# Student Survey Questions

<table>
<thead>
<tr>
<th>Question- The 2-stage Exam......Group part....</th>
<th>Agree</th>
<th>Disagree</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provided a learning opportunity which enhanced my knowledge &amp; understanding of the subject matter</td>
<td>88%</td>
<td>12%</td>
<td></td>
</tr>
</tbody>
</table>

Rate the help-fullness of this 2-stage exam towards aiding your understanding of **Relative & Absolute Time**

<table>
<thead>
<tr>
<th>Least Helpful</th>
<th>Relative &amp; Absolute Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Relative</td>
</tr>
<tr>
<td>2</td>
<td>Absolute</td>
</tr>
<tr>
<td>3</td>
<td>Relative</td>
</tr>
<tr>
<td>4</td>
<td>Relative</td>
</tr>
<tr>
<td>5</td>
<td>Absolute</td>
</tr>
<tr>
<td>6</td>
<td>Absolute</td>
</tr>
<tr>
<td>7</td>
<td>Relative</td>
</tr>
<tr>
<td>8, 9, 10</td>
<td>Most Helpful</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Ranking</th>
<th>Students (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>47%</td>
</tr>
<tr>
<td>9</td>
<td>35%</td>
</tr>
<tr>
<td>8</td>
<td>18%</td>
</tr>
</tbody>
</table>

Most helpful for learning

Least helpful for learning

---

Data_plots vs % of students Montoya survey

B-% students

A-rating

Data_plots vs % of students Montoya survey

B-% students

A-rating
Summary and Conclusions

- Designed, conducted, evaluated a 2-stage exam pedagogy for an introductory physical geology class

- Analysis of the trends in student performance for Individual-Group test components
  - Independent of group composition most all students improved scores in the collaborative part of the exam
  - Statistically significant difference in student performance → improved group exam

- Analysis of trends in student performance-level with respect to group test scores
  - High performers lower benefit from group exam (indiv. to group exam) (Improved mean score by 11% for both exams)
  - Low & Mid level Performers largest benefit from group exam T#3 Improved mean score by 89% 62%, T #4 by 44.6%, 42%

- Analysis of student learning from group exam component
  - Statistical difference between Individual & Learning scores (means)
  - All students regardless of performance-level benefited from 2-stage exam
  - High performers benefited the least (Mean score improved by 11% on Retest)
  - Mid & Low performers benefited the most (Mean score improved by ~16%/Mid & 16.6%/Low)

- 2-Stage exams can be an assessment method + a learning tool
  - Learning is most effective after students have studied the material
  - Students learn by both teaching & learning from one another
  - Students & Instructors both benefit
Where To Go → Next Steps

1. Collect more data → more experiments

2. Test question design emphasize evaluation & analysis

3. Understand the repeat test effect
   Differentiate learning effects as a function of “repeat test” and “collaboration”
   
   *Did students do better on learning test because they saw the test before or was it the result of learning from collaboration*

4. Understand effect of student group dynamics on student learning
   Random grouping vs. engineered grouping methods

5. Determine the agility/robustness of the 2-stage test methodology
   Examine effects across other courses & instructors
Final Question

Who Wants to do 2-Stage Exam Again...Show of Hands!

Thank You For Your Attention!
End of Presentation
New Mexico, the struggle of student success is perpetual

- state’s high poverty rate
- low graduation rate
- NM students failing to meet goals in reading and math.
- According to Education Week magazine’s latest “Quality Counts” study New Mexico ranked 49th (out of the 50 states and District of Columbia) for educational quality in a national review of measures such as high school graduation rates, advanced placement exam results, school finances and pre-K enrollment.
Summary and Conclusions + Next Steps

• Higher administrative efforts
• Group composition differences may limit effectiveness of approach (dominant student or groups with free-riders)
## Statistical Analysis for test # 3 (n=42) & #4 (n=40)

### Individual and Group Test Scores

<table>
<thead>
<tr>
<th>Values</th>
<th>Test #3 (n=42)</th>
<th>Test #4 (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Exam</td>
<td>Group Exam</td>
</tr>
<tr>
<td>Minimum</td>
<td>36</td>
<td>60</td>
</tr>
<tr>
<td>Maximum</td>
<td>84</td>
<td>96</td>
</tr>
<tr>
<td>Mean</td>
<td>56</td>
<td>79.52</td>
</tr>
<tr>
<td>Median</td>
<td>56</td>
<td>76</td>
</tr>
<tr>
<td>RMS</td>
<td>58.018</td>
<td>80.207</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.3528</td>
<td>10.572</td>
</tr>
<tr>
<td>Variance</td>
<td>235.71</td>
<td>111.77</td>
</tr>
<tr>
<td>Stnd. Error</td>
<td>2.369</td>
<td>1.6313</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.33792</td>
<td>-0.054341</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-1.0392</td>
<td>-1.112</td>
</tr>
</tbody>
</table>
### Results: Paired t-Test for test #3 (n=42) & #4 (n=40)

*Paired t-Test – compare means on the same subjects over time or under different conditions*

<table>
<thead>
<tr>
<th></th>
<th>Test #3 (n=42)</th>
<th></th>
<th>Test #4 (n=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Exam</td>
<td>Group Exam</td>
<td>Individual Exam</td>
</tr>
<tr>
<td>Mean</td>
<td>56</td>
<td>79.52</td>
<td>56.85</td>
</tr>
<tr>
<td>Variance</td>
<td>235.707</td>
<td>111.768</td>
<td>271.5667</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.3528</td>
<td>10.572</td>
<td>16.4793</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>2.36898</td>
<td>1.6313</td>
<td>2.6056</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-23.5238</td>
<td></td>
<td>-15.45</td>
</tr>
<tr>
<td>t-value</td>
<td>-10.459</td>
<td></td>
<td>-7.277</td>
</tr>
<tr>
<td><strong>t-Probability</strong></td>
<td><strong>&lt;0.0001</strong></td>
<td></td>
<td><strong>&lt; 0.0001</strong></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.41595</td>
<td></td>
<td>0.68039</td>
</tr>
<tr>
<td>Corr. Probability</td>
<td>0.00615</td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

The t-probability value determines if there is a statistically significant difference between the two means. If this value is below a certain level (~0.05) the conclusion is that there is a difference between the two group means. A t-probability <0.0001 → the probability of the two scores being the same is low (99.99% that the two scores are different).
A Closer Look at the Data – Box Plots

- Box Plot quickly examine data and extract their major characteristics
- Global behavior
- Non-parametric (display variation in samples w/o making any assumptions of underlying stats)
  Does not tell you the distribution formula, due not give you an indication of formula for the distribution
  i.e., Gausian has 2 parameters to define it (mean & std dev) box plot does not provide the distribution
- Spacing between different parts of box →
  degree of dispersion (spread) & skewness in data

Box represents each variable as a separate box

Y axis → range of data (scores)

Maximum Value 95% of population/Scores below this pt. (2 sigma)

Upper Quartile 75% of population/scores below this pt.

Median value

Lower Quartile 25% of population/scores below this pt.

Minimum Value 5% of population/scores below this pt (2 sigma)

Outliers (values outside the range)
Influence of Student Performance Level on Group Exam Scores

- **High performers** least difference between individual and group exam scores
- **High performers** lower benefit from group exam
  - More students scores below median value (skewed)
  - Moderate degree of dispersion (spread) in data (moderate box size)

---

*Box Plots - Test #3*

- **Least Benefit**
  - High Performers
  - Mid Performers
  - Low Performers

---

Reference slide save
Influence of Student Performance Level on Group Exam Scores

- High performers least difference between individual and group exam scores
- High performers lower benefit from group exam
  - More students scores below median value (skewed)
  - Moderate degree of dispersion (spread) in data (moderate box size)
- Low Performers largest difference between individual and group exam scores
  - Most benefit from group exam
  - Lowest degree of dispersion within test type data (narrow boxes)
- Mid Level performers close to that of low performers
### Student Benefit by Performance Level → Box Plot Test #3 (7-7-17)

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv.</th>
<th>Group</th>
<th>Mean</th>
<th>Indiv.</th>
<th>Group</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75.6</td>
<td>84</td>
<td>49.2</td>
<td>79.8</td>
<td>74.57</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>76</td>
<td>88</td>
<td>52</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

- High performers: highest mean and median scores vs Mid & Low performers

**Least Benefit**

- High Performers

**Most Benefit**

- Mid Performers

- Low Performers

Reference slide save
Student Benefit by Performance Level → Box Plot Test #3 (7-7-17)

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv.</th>
<th>Group</th>
<th>Indiv.</th>
<th>Group</th>
<th>Indiv.</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>75.6</td>
<td>84</td>
<td>49.2</td>
<td>79.8</td>
<td>39.4</td>
<td>74.57</td>
</tr>
<tr>
<td>Median</td>
<td>76</td>
<td>88</td>
<td>52</td>
<td>76</td>
<td>38</td>
<td>76</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>7.14</td>
<td>9.23</td>
<td>8.97</td>
<td>10.26</td>
<td>3.797</td>
<td>8.39</td>
</tr>
<tr>
<td>Min Score</td>
<td>64</td>
<td>68</td>
<td>36</td>
<td>60</td>
<td>36</td>
<td>60</td>
</tr>
<tr>
<td>Max Score</td>
<td>84</td>
<td>96</td>
<td>60</td>
<td>96</td>
<td>44</td>
<td>88</td>
</tr>
</tbody>
</table>

- High performers highest mean and median scores vs Mid & Low performers

Reference slide save
Individual vs. Group Scores (median/mean) as a function of student performance

- **High performers** least difference in median (mean) scores between exams
  - More students scores below median value (skewed)
  - Moderate degree of dispersion (spread) in data

- Mid & low performers higher difference in median (mean) scores
  - Both have larger dispersion with in group scores (Group Effect)
  - Both have minimal dispersion in the individual scores
Student Benefit by Performance Level → Box Plot Test #4 (7-7-17)

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv.</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>74.5</td>
<td>82.6</td>
</tr>
<tr>
<td>Median</td>
<td>74</td>
<td>78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv.</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>51.4</td>
<td>73</td>
</tr>
<tr>
<td>Median</td>
<td>51</td>
<td>71</td>
</tr>
</tbody>
</table>

Reference slide save
Reference slide save
Paired t-Test for test # 4 individual & Retest
Individual and Retest Scores

Paired t-Test –

✓ Compare means on the same subjects over time or under different conditions
✓ Used to determine if two data sets are significantly different from each other

<table>
<thead>
<tr>
<th>Variance</th>
<th>252.975</th>
<th>255.139</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Dev.</td>
<td>15.9052</td>
<td>15.9731</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>2.68847</td>
<td>2.69994</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-7.2</td>
<td></td>
</tr>
<tr>
<td>t-value</td>
<td>-7.7256</td>
<td></td>
</tr>
<tr>
<td><strong>t-Probability</strong></td>
<td><strong>&lt;0.0001</strong></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.94018</td>
<td></td>
</tr>
<tr>
<td>Corr. Probability</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

T-test

Determine if there is a statistical difference between Individual & restest scores (means)
t-probability >0.0001
T-test results demonstrated that the results are statistically different
There is a statistically significant improvement in student performance between individual and the learning test
Ind test serves as baseline assessment of students knowledge
Improvement on leaning test indicates
a) During the collaborative test the students acquitted knowledge they did not have previously
b) Learning was sufficient enough for students to remember it 56 weeks later
### Paired t-Test for test #4 individual & Retest Scores

<table>
<thead>
<tr>
<th></th>
<th>Test #4 (n-35)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Exam</td>
<td>Retest Exam</td>
</tr>
<tr>
<td>Mean</td>
<td>55.71</td>
<td>62.91</td>
</tr>
<tr>
<td>Variance</td>
<td>252.975</td>
<td>255.139</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.9052</td>
<td>15.9731</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>2.68847</td>
<td>2.69994</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-7.2</td>
<td></td>
</tr>
<tr>
<td>t-value</td>
<td>-7.7256</td>
<td></td>
</tr>
<tr>
<td>t-Probability</td>
<td><strong>&lt;0.0001</strong></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.94018</td>
<td></td>
</tr>
<tr>
<td>Corr. Probability</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

T probability value determines if there is a statistically significant difference between the two means. If this value is below a certain level (~0.05) the conclusion is that there is a difference between the two group means.

T-probability <0.0001 → the probability of the two scores being the same is low (99.99 % that the two scores are different)
Student Benefit by Performance (Test #4) (7-7-17)

<table>
<thead>
<tr>
<th>Exam Type</th>
<th>Indiv.</th>
<th>Group</th>
<th>R-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>74.5</td>
<td>82.6</td>
<td>80.3</td>
</tr>
<tr>
<td>Median</td>
<td>74</td>
<td>78</td>
<td>83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exam Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

**Exam Score (%)**

- **High Performers**
  - Least difference between Individual and Retest exam scores
  - Moderate degree of dispersion (spread) in data
  - Mean score improved by 11% on Retest

- **Mid Performers**
  - Moderate degree of dispersion (spread) in data
  - Mean score improved by ~16% (Mid) & 16.6% (Low)

- **Low Performers**
  - Least difference between Individual and Retest exam scores
  - Low degree of dispersion in data
  - Mean score improved by ~16% (Mid) & 16.6% (Low)

- **Methods to improve Low and Mid performers learning are critical in higher education**
  - Benefit directly result of peer teaching
  - Benefit due to greater motivation to learn material (better attitude about learning)
  - Student better prepared- “desire to look good amongst peers”