10 Reasons to Replace Midterms with Weekly Online Quizzes

Dave Pritchard, Cecil and Ida Green Prof. Physics
Byron Drury, Postdoc in RELATE
REsearch in Learning Assessing and Tutoring Effectively
1. Evidence Midweek Testing > Midweek Review

- Learn Monday, Test Friday
- Wednesday has **Review and Test**, some words omitted
- Clearly Wednesday testing vs not testing determines Friday test score

Karpicke & Roediger

*Science* 319, 966 (2008)

**Fig. 2.** Proportion recalled on the final test 1 week after learning. Error bars represent standard errors of the mean.
When do Students Read e-text; traditional?

Is this what we want? Changing Studying Behavior!

→ Write Handbooks, not Textbooks
Question? “How often do you Study, by which I mean review & reflect & do extra problems vs. just doing the weekly p-sets?"

“I just study for exams”

“Only if I have a test”

“Are you kidding?”
2 text reading depends on Course Structure

**Traditional**

Traditional: Lectures, etext, weekly homework

**Flipped**

Flipped: Flipped Classroom with preclass HW

**Reformed & Biweekly**

Reformed & Biweekly: Same plus quizzes

Analyzing the Impact of Course Structure on Electronic Textbook Use in Blended Introductory Physics Courses
Daniel Seaton..DEP Am. J. Phys. 82, 1186 (2014); http://dx.doi.org/10.1119/1.4901189
Net of 1.1 std dev for about 60% copying implies ~ 2.0 std. dev. effect size for no copying vs. all copying.

One might expect that *decreasing* the use of unsupervised answer-sharing forums would improve final exam scores.

The equation for predicting final exam scores is:

\[
\text{final exam score} = -0.47C + 0.26EX1 + 0.26S + 0.20D
\]

- $C$: Copying online HW
- $EX1$: First Midterm EXam
- $S$: Score on HW
- $D$: Diagnostic Test
3. Monitored vs. Unmonitored Discussion Forum

- Students switched to using discussion forum monitored by TA’s vs. ones that allow posting of answers

Laverty et. al. *Want to Reduce Guessing and Cheating While Making Students Happier? Give More Exams!*
Weekly Quizzing Should Increase Learning

Summary

• Rodighier: Quizzing Improves Learning
• Students Study Every Week (not just do HW)
• Students don’t use homework copying forums

Indeed, it Dramatically Increases Final

• Final average shifted from ~43% to ~68%.
• Large reduction in D & F grades

Fig. 6. Final exam grade distribution.
4. Students liked frequent quizzes

[Students] “seemed unhappy with the [testing] system when they first heard about it”, but gave very much more positive opinions after trying it.

Fig. 5. Student satisfaction with exam schedule.
Gave weekly quizzes in Friday’s class

½ online and ½ on-paper

Students could allocate time as they desired;

online/on-paper correlation: $r \approx 0.3$ (week-by-week)

$r \approx 0.7$ (average over weeks)
Online Quiz: Concept vs. Calculation

A blend of question types and response formats:

“conceptual:” multiple choice, checkbox
“calculational:” symbolic input, numerical

Average ~3 min/question, but with large variation
Conceptual shorter, Calculational longer
~10 questions for ½ hour quiz

Administered in-class

Students bring computers, ODL supplies loaners (~6%)
Quiz Grading

Many online questions allow multiple attempts;

We can measure student skill better if we give partial credit for correct answers submitted after initial incorrect attempts.

Different problem types give different information about student knowledge.

We explored a family of grading schemes:

\[ s_{adj}(\theta, P_M, P_O) = \cos^2(\theta) [(1 - P_M) s_{M,1st} + P_M s_{M,ev.}] + \sin^2(\theta) [(1 - P_O) s_{O,1st} + P_O s_{O,ev.}] \]

\( \theta \) controls relative weight of MC and OR questions.

\( P_{M/O} \) are amount of credit awarded for correct answers given \textit{after} an incorrect attempt on MC/OR.
Test Reliability

\[ \rho = \frac{\sigma_T^2}{\sigma_O^2} \]

\[ \sigma_O^2 = \sigma_T^2 + \sigma_E^2 \]
Quiz Week to Week Self-Consistency

Need to quantify effects of testing error
use z-scores to normalize for week-to-week variation in difficulty

\[ z = \frac{\text{score} - \text{quiz mean}}{\sigma_{\text{quiz}}} \]

Quantify the self consistency with Cronbach’s \( \alpha \)

\[ \alpha = \frac{K \bar{c}}{\bar{v} + (K - 1) \bar{c}} \]

\( K \) is # of tests, \( \bar{v} \) is mean test variance, \( \bar{c} \) is mean test covariance

Quantifies the extent to which the quizzes measure the same underlying construct
Ranges from 0 (no correlation) to 1 (perfect correlation)
Rule of thumb: 0.7+ is ‘acceptable,’ 0.8+ is ‘good’
Quiz Correlation with Final

Partial Credit Awarded for Correct Answer After Incorrect First Attempt
Different Question Formats test Different Knowledge?

![Graph showing correlation between question formats and knowledge level](image-url)
Online quizzes are more consistent than written and predict performance on the final better

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Cronbach’s $\alpha$</th>
<th>Corr. w/ Final</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written</td>
<td>0.65</td>
<td>0.76</td>
<td>0.00110</td>
</tr>
<tr>
<td>Online First attempt</td>
<td>0.74</td>
<td>0.83</td>
<td>0.00012</td>
</tr>
<tr>
<td>70% credit</td>
<td>0.78</td>
<td>0.86</td>
<td>0.00004</td>
</tr>
<tr>
<td>Any attempt</td>
<td>0.76</td>
<td>0.86</td>
<td>0.00005</td>
</tr>
<tr>
<td>Optimized</td>
<td>0.80</td>
<td>0.88</td>
<td>0.00002</td>
</tr>
</tbody>
</table>
Weekly Online Quizzing Summarized

Fairer, and predicts every part of Final better than rubric-graded symbolic response quizzes, eliminates grading error

Facilitates weekly quizzes: research → less dishonesty, far better learning

Less faculty time spent on grading/testing

Facilitates education research

Year-to-year and student-to-student grading consistency → evaluate teaching by results
Discussion

• Difficulties with student acceptance?

• Quiz Security vs. Releasing Quizzes
  • Is knowing in real time enough?
  • Better than paper quizzes in fraternity Bibles?
  • Online offers easy checks

• Implementing in multi-section TEAL subjects?