Beyond the Foundations

A quantitative investigation of Cerego's impact on knowledge transfer and understanding

Iain M. Harlow

Students at the University of Hawaii showed a 12-point improvement in final exam score per level of retention built in Cerego - including a larger 16-point improvement per level on questions designed to test higher level understanding and analytical skills.

Contents

1 Introduction

As a learning platform used and trusted across important domains, from higher education to casualty care training, Cerego's most important objective is to help learners build real, lasting expertise in whatever they need to learn. To do so successfully requires three things:

Retention: Learners remember more, longer

Knowledge Transfer: Learners can apply knowledge outside the system

Understanding: Learners develop understanding, not simply a list of facts

There is substantial evidence that concepts learned in Cerego are remembered for longer. Within Cerego, we can track in detail how knowledge retention improves as learners complete reviews. For example, when learners complete four reviews of a concept in Cerego - roughly two minutes of interaction spread across a week - their knowledge for that concept sticks around for weeks after. Accuracy on such concepts is maintained at around 90% for several weeks, compared to 60% for concepts studied but not reviewed.

Because Cerego tracks the predicted accuracy for every concept being learned, by every user and at every moment in time, we have a detailed and constantly updating model of how concept difficulty, learner ability and past learning history contribute to memory retention, as well as an unprecedented insight into the knowledge state of each learner and concept at any given moment. But how transferable is this knowledge? That is, to what extent does learning in Cerego result in improved performance outside the system - in a different, real-world context?

2 Knowledge Transfer: Cerego effect on Exam Scores

Recently, we were able to test the transfer of knowledge using a more detailed dataset. As part of the Next Generation Courseware Challenge (a Bill & Melinda Gates Foundation initiative), Cerego partnered with higher education institutions to improve the platform and measure its efficacy, with a particular focus on at-risk groups such as Pell grant eligible or first generation college students.

At one institution in particular, the University of Hawai'i at Manoa, we were able to gather detailed anonymized data from a Introduction to Macroeconomics course of 98 students. The data included measures of prior educational attainment (grade point average, GPA, at the beginning of the course), Pell grant eligibility, and performance on the midterms and final exam. Most interestingly, the final exam was divided into two main sections: the first tested knowledge directly using multiple choice questions, and the second section tested the application of principles through two longer, more discursive questions. This allowed us to separately examine whether Cerego use was associated with better than expected performance on multiple choice questions, discussion questions, or both.

Our first analysis was a simple one - after controlling for the students' expected level of performance (using prior GPA) as well as demographics like age, gender and Pell status, did the amount of retention built using Cerego predict overachievement on the final exam?

In short, yes - substantially. A multiple regression analysis (Figure 1) with factors of Cerego retention (in average level per item, range 0-2.75, average 0.75) and prior GPA (range 0-4, average 2.90) found both to be strongly and separately predictive of final exam score. Each point of prior GPA was associated with a 13.5 percentage point increase in exam score, while each level of retention built in Cerego by the date of the exam was associated with a further **12 point increase**: more than a full letter grade.

Follow-up two-way ANOVA was used to check the robustness of the findings and test for interactions. The ANOVA yielded main effects for both GPA [$F(1, 94) = 31.224, p < .001$] and Cerego retention [$F(1, 94) = 11.215, p = .001$] but no interaction, suggesting Cerego retention improved exam scores for students with both low and high educational attainment.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   exam   R-squared:                       0.499
Model:                            OLS   Adj. R-squared:                  0.489
Method:                 Least Squares   F-statistic:                     47.34
Date:                Tue, 27 Feb 2018   Prob (F-statistic):           5.45e-15
Time:                        13:27:30   Log-Likelihood:                 43.701
No. Observations:                  98   AIC:                            -81.40
Df Residuals:                      95   BIC:                            -73.65
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1372      0.061      2.238      0.028       0.015       0.259
gpa            0.1350      0.024      5.590      0.000       0.087       0.183
cerego         0.1169      0.035      3.350      0.001       0.048       0.186
==============================================================================
Omnibus:                       14.857   Durbin-Watson:                   2.462
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               20.742
Skew:                          -0.715   Prob(JB):                     3.13e-05
Kurtosis:                       4.743   Cond. No.                         13.6
==============================================================================
```

Figure 1: Regression model predicting exam score from prior grade point average (*gpa*), retention level built (*cerego*), and a constant (*const*). More complex models including demographic factors did not improve the model fit and gave similar results for GPA and Cerego retention. Standardizing variables or log-transforming the Cerego variable also led to the same conclusion: Cerego retention predicted significantly higher exam performance even after accounting for prior educational attainment.

These results are encouraging for three main reasons. First, they demonstrate a strong link between retention built in Cerego, and a real-world outcome (in this case a college exam score). Second, this link was present even after accounting for prior educational attainment in the form of GPA. Finally, the link existed for students of both low and high GPA scores: studying in Cerego was beneficial regardless of the level of ability previously demonstrated in other courses.

3 Building on the Foundation: Cerego effect on Understanding

Retention built in Cerego was followed by overachievement on the final exam for this course, mirroring similar evidence of knowledge transfer in high school students in Florida, dental school students at NYU, and Syrian refugees in Turkey.

The most interesting aspect of this dataset, though, is the ability to look at performance on individual questions - which lets us examine whether Cerego retention helped only with foundational knowledge (as measured by multiple choice questions), or whether it also improved understanding and application (as measured by open-ended analytical or discursive questions).

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 exam_mc   R-squared:                       0.442
Model:                             OLS   Adj. R-squared:                  0.430
Method:                  Least Squares   F-statistic:                     37.57
Date:                 Mon, 05 Mar 2018   Prob (F-statistic):           9.53e-13
Time:                         12:24:05   Log-Likelihood:                 43.450
No. Observations:                   98   AIC:                            -80.90
Df Residuals:                       95   BIC:                            -73.14
Df Model:                            2
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.2407      0.061      3.917      0.000       0.119       0.363
gpa            0.1230      0.024      5.077      0.000       0.075       0.171
cerego         0.1005      0.035      2.873      0.005       0.031       0.170
==============================================================================
Omnibus:                       34.523   Durbin-Watson:                   2.264
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               81.731
Skew:                          -1.289   Prob(JB):                     1.79e-18
Kurtosis:                       6.657   Cond. No.                         13.6
==============================================================================
```

Figure 2: Regression model predicting percentage score on one-point multiple choice questions (*exam_mc*) as a function of prior grade point average (*gpa*), average retention level built in Cerego (*cerego*), and a constant (*const*). Both predictors were significant in follow-up ANOVA analyses.

To begin to answer this, we applied the same multiple regression analysis as in the previous section, but this time instead of predicting exam performance overall we separately predicted performance on multiple choice questions (Figure 2) or on analytical questions (Figure 3). Cerego retention predicts overperfomance on the

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              exam_disc   R-squared:                       0.341
Model:                            OLS   Adj. R-squared:                  0.327
Method:                 Least Squares   F-statistic:                     24.56
Date:                Mon, 05 Mar 2018   Prob (F-statistic):           2.52e-09
Time:                        12:23:27   Log-Likelihood:                -13.751
No. Observations:                  98   AIC:                             33.50
Df Residuals:                      95   BIC:                             41.26
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1477      0.110     -1.340      0.183      -0.366       0.071
gpa            0.1683      0.043      3.876      0.000       0.082       0.254
cerego         0.1620      0.063      2.584      0.011       0.038       0.287
==============================================================================
Omnibus:                        2.509   Durbin-Watson:                   2.300
Prob(Omnibus):                  0.285   Jarque-Bera (JB):                2.523
Skew:                           0.361   Prob(JB):                        0.283
Kurtosis:                       2.690   Cond. No.                         13.6
==============================================================================
```

Figure 3: Regression model predicting percentage score on ten-point analytical and discussion questions (*exam_disc*) as a function of prior grade point average (*gpa*), average retention level built in Cerego (*cerego*), and a constant (*const*). Both predictors were significant in follow-up ANOVA analyses.

exam relative to that expected from prior GPA - does this exist only for multiple choice questions, or also for understanding-focused questions?

In fact, Cerego retention was significantly and positively associated with over-performance on both sections of the exam. Far from only improving foundational knowledge, Cerego retention had a numerically stronger link with over-performance on the higher-order analytical questions than the multiple choice questions: after accounting for GPA, Cerego retention was associated with a 10 percentage point increase per level on multiple choice performance, but a larger **16 point increase** on discussion question performance.

As a direct test of whether Cerego retention built higher order analytical skills, we constructed a further multiple regression model predicting exam discussion question performance using Cerego retention and the multiple choice performance from the same exam. In other words, instead of accounting for general student ability as measured by GPA, this model factored out foundational level knowledge that was specifically demonstrated during the exam itself.

Supposing we already know the scores on the multiple choice section of the exam, to what extent does the Cerego retention a student built up over the course tell us anything extra about their understanding? If Cerego only builds up foundational knowledge, there should be no relationship (or even a negative one, if Cerego helps on the foundational knowledge section more than the higher order questions).

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              exam_disc   R-squared:                       0.384
Model:                            OLS   Adj. R-squared:                  0.371
Method:                 Least Squares   F-statistic:                     29.61
Date:                Mon, 05 Mar 2018   Prob (F-statistic):           1.01e-10
Time:                        17:02:09   Log-Likelihood:                -10.433
No. Observations:                  98   AIC:                             26.87
Df Residuals:                      95   BIC:                             34.62
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1555      0.094     -1.647      0.103      -0.343       0.032
exam_mc        0.7518      0.158      4.768      0.000       0.439       1.065
cerego         0.1490      0.059      2.522      0.013       0.032       0.266
==============================================================================
Omnibus:                        3.204   Durbin-Watson:                   2.024
Prob(Omnibus):                  0.201   Jarque-Bera (JB):                2.117
Skew:                           0.153   Prob(JB):                        0.347
Kurtosis:                       2.348   Cond. No.                         9.64
==============================================================================
```

Figure 4: Regression model predicting percentage score on ten-point analytical and discussion questions (exam_disc) as a function of multiple choice section performance on the same exam ($exam_mc$), average retention level built in Cerego (*cerego*), and a constant (*const*). Both predictors were significant in follow-up ANOVA analyses.

In fact, we see a strong overperformance on this section of the exam for the students who built up their Cerego retention, of almost **15 percentage points** per level. That is, even when we have up to date knowledge of what students know, from the multiple choice section of the very exam we are analyzing, their Cerego retention tells us considerably *more* about how much they actually understand and can apply. Faced with two students with identical scores on a multiple choice knowledge test, the student with the higher Cerego retention is more likely to have a deeper understanding of the material.

4 How does Cerego help build Understanding?

There are many reasons why Cerego might specifically help to build understanding, rather than only foundational knowledge. One interesting possibility is that knowledge and expertise compound. When a learner can more readily bring to mind the foundational terminology and principles that underpin a subject, they are better able to follow, understand, and incorporate new information about the subject.

The improvement in understanding we measured for Cerego learners may be indirect evidence for this effect - learners who had built more Cerego retention displayed better understanding on the final exam. We also, however, tested the effect more directly, by surveying learners on their ability to follow the lectures during the course. Using the same modeling approach as in previous analyses, we found that after controlling for GPA, learners who built more Cerego retention during the course reported being better able to follow the lectures.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     survey_follow_lectures   R-squared:                       0.093
Model:                                OLS   Adj. R-squared:                  0.061
Method:                     Least Squares   F-statistic:                     2.965
Date:                    Wed, 21 Mar 2018   Prob (F-statistic):             0.0594
Time:                            09:10:57   Log-Likelihood:                -77.859
No. Observations:                      61   AIC:                             161.7
Df Residuals:                          58   BIC:                             168.1
Df Model:                               2
Covariance Type:                nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.0830      0.488      8.372      0.000       3.107       5.059
gpa           -0.3951      0.188     -2.098      0.040      -0.772      -0.018
cerego         0.5767      0.259      2.228      0.030       0.059       1.095
==============================================================================
Omnibus:                        0.048   Durbin-Watson:                   2.176
Prob(Omnibus):                  0.976   Jarque-Bera (JB):                0.225
Skew:                           0.024   Prob(JB):                        0.894
Kurtosis:                       2.706   Cond. No.                         15.6
==============================================================================
```

Figure 5: Regression model with factors of prior GPA and Cerego retention, predicting response on a 5-point Likert scale to the question "How well did you manage to follow the lectures in this course, compared to other courses?". Both predictors, but not their interaction, were significant in follow-up ANOVA analyses.

These survey data are significant, since they get more directly at a potential causal mechanism for Cerego's effect on understanding: Cerego equipped learners to better understand and integrate new information as it was presented to them.

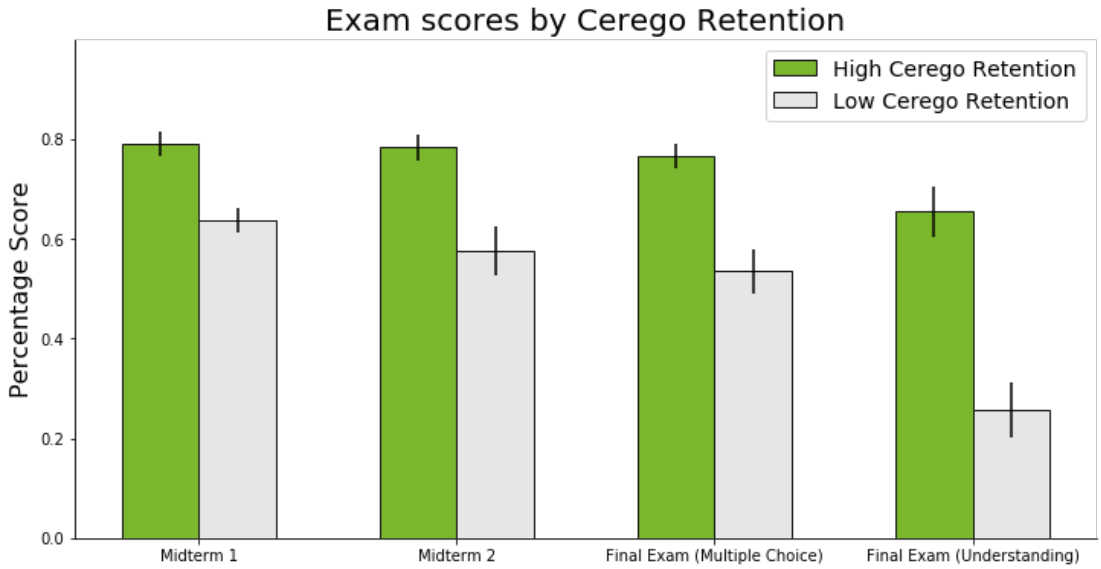

Figure 6: Scores on two midterm exams and one final exam reveal a substantial and increasing gap in educational attainment between students who built high (top third) and low (bottom third) Cerego retention. The visible trends of greater exam performance by students who built high Cerego retention, and the increase in this difference from early midterms to the final exam, were both statistically significant. Two-way ANOVA revealed a significant main effect for Cerego retention ($p < .001$), as well as a significant interaction ($p = .040$) between Cerego retention and exam order which reflected an increasing attainment gap over time.

Corroborating the theory that Cerego improved understanding by helping users engage with the material more deeply as it was taught, the gap in exam achievement between high and low Cerego users widened as the course progressed (Figure 6). Performance gaps between the top and bottom third of learners by Cerego retention grew from 15 percentage points on Midterm 1, to 21 points on Midterm 2, and reached 28 by the final exam (including a 40 percentage point difference on the final two exam questions, which focused on understanding).

A similar pattern was observed in a series of 8 quizzes administered in class throughout the course: an average 17 point difference over the first two quizzes grew to a 38 point difference by the final two quizzes (Figure 7).

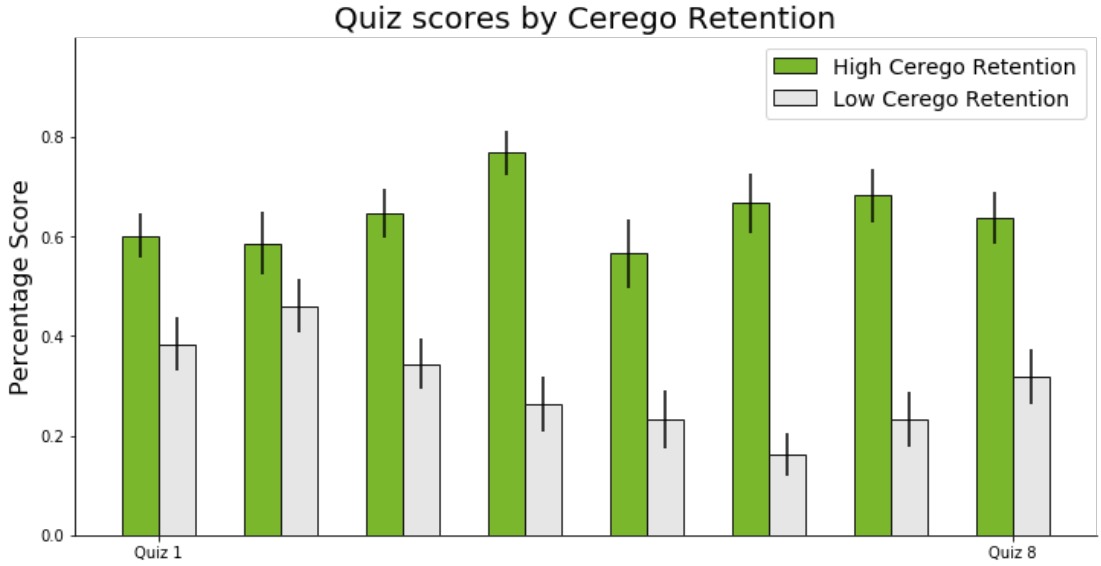

Figure 7: Scores on graded, in-class quizzes throughout the course reveal a substantial and increasing gap in educational attainment between students who built high (top third) and low (bottom third) Cerego retention. The visible trends of greater quiz performance by students who built high Cerego retention, and the increase in this difference during the course, were both statistically significant. Two-way ANOVA revealed a significant main effect for Cerego retention ($p < .001$), as well as a significant interaction ($p = .003$) between Cerego retention and quiz order which reflected an increasing attainment gap over time.

These trends, reflected in statistically significant interaction terms between Cerego and exam/quiz order, are important for two reasons. First, they support the theory advanced above that Cerego-related improvement in final exam score may be mediated by better ongoing understanding of the course material. In other words, higher Cerego retention reflects a better mastery of the material, which in turn allows deeper ongoing engagement with the course as more advanced concepts are introduced - reflected here in the widening attainment gap over time, as well as survey data reporting better ability to follow lectures in students with greater Cerego retention.

Secondly, they bolster the conclusion from earlier analyses that Cerego retention is not simply correlated with exam performance because the more able students used Cerego more - if this were the case, the attainment gaps should be steady throughout the course. The fact that they increase over time from around a 15%

gap to as much as 40% strongly suggests that Cerego retention was not a symptom but a direct cause of student success in the course.

Overall, the data from this study provide novel evidence to support the view that developing mastery over the foundational content of a subject - as Cerego is designed to do - is not equivalent to a rote learning of isolated facts. Instead, this foundational mastery is critical for developing the deeper understanding and expertise that we generally desire from a high quality course.

5 Appendix

5.1 Data sample

The dataset consists of 98 anonymous college students (47 female, mean age 20.8, range 17-41) at the the University of Hawai'i, Manoa, who were assigned Cerego as part of a 16-week Macroeconomics course. Students were provided with Cerego courseware consisting of 19 assignments (770 items) and were given a goal of reaching Level 1 retention for each. Level 1 retention in Cerego generally takes 3-4 interactions per item to achieve (around 2 minutes of interaction per item, split across 4-6 days on average).

Cerego use was graded, with 10% of the course grade determined by progress students made towards this goal. Mean Cerego retention level reached was 0.75. No analyses in the report use course grade as an outcome variable since it is not independent of Cerego retention; instead we analyze externally graded outcomes such as in-class quizzes, formal midterms and the final exam.

5.2 Data Privacy

Data were collected as part of the Next Generation Courseware Challenge, overseen by SRI and funded by the Bill & Melinda Gates Foundation.

The study was approved by the University of Hawaii at Manoa IRB (Protocol ID: 2016-30534). Cerego take student data privacy extremely seriously and took additional precautions to ensure that at no stage were Cerego, or any party outside of the UH Manoa Dean's Office, in possession of personally identifiable outcome (e.g. grade) or demographic data. All Cerego systems data and survey responses used in the study were securely provided to the participating institution (UH Manoa) where they were combined with outcome and demographic data and anonymized. This final, anonymized dataset was shared to a secure SRI server before being used to create the report above, and contains no names, email addresses, user id or other identifying information.

5.3 Survey Data

Ten survey questions were asked of learners at the end of the course; six to measure the students' own sense of how useful Cerego was, one baseline measure of perceived course difficulty as a control, and three designed to more closely examine how Cerego affected the way learners studied. All of the responses were gathered using a 5-point Likert scale, with 5 indicating strongest agreement with the statement or most strongly positive response to the question, 1 indicating strongest disagreement, and a neutral response at position 3. Figure 8 shows the question wording and a visual summary of the responses for each.

Of the 98 students offered the survey by email, 61 responded. The respondents were a generally representative sample in terms of ability: a small difference existed in prior GPA (3.02 for respondents vs 2.70 for non-respondents, $p = .056$) but none in performance on the first in-class quiz (49% v 47%, $p = .799$) or the first midterm exam (73% v 71%, $p = .704$), and average survey ratings for difficulty of the course were very close to neutral (33% said the material was "more difficult" than other courses, 33% said "less difficult").

Overall response to Cerego was good, and trended more towards appreciation of the effectiveness than simply enjoyment of the platform. Of the four simple value assessments, 72% agreed or strongly agreed (v 11% disagreed) with the statement "I like learning with Cerego", 90% v 8% agreed that "Cerego improved my knowledge of the subject", 79% v 11% agreed that "Cerego helped (or will help me) get a better grade in the class" and 54% v 20% agreed with the statement "I would like to use Cerego in my future classes".

On the specific effectiveness of Cerego at building foundational knowledge that lasts, majorities of students also felt that the system worked as designed: 51% v 11% agreed that their retention for material learned through Cerego was better than in comparable classes, and 67% v 3% reported that they "learned the foundational content of this course" better than they did for other courses.

There was little evidence that positive reactions to Cerego were more likely to come from higher-ability students (mean effect on Likert scale per point of prior GPA was +0.13, with a mixture of positive and negative trends; p-values starting at 0.136 and averaging 0.453). Interestingly, there was some evidence that Pell-eligible students appreciated Cerego at least as much and possibly more than other students: multiple regression with factors of GPA and Pell-eligibility for the six valuation questions showed no effects of GPA, but a trend in every case (average

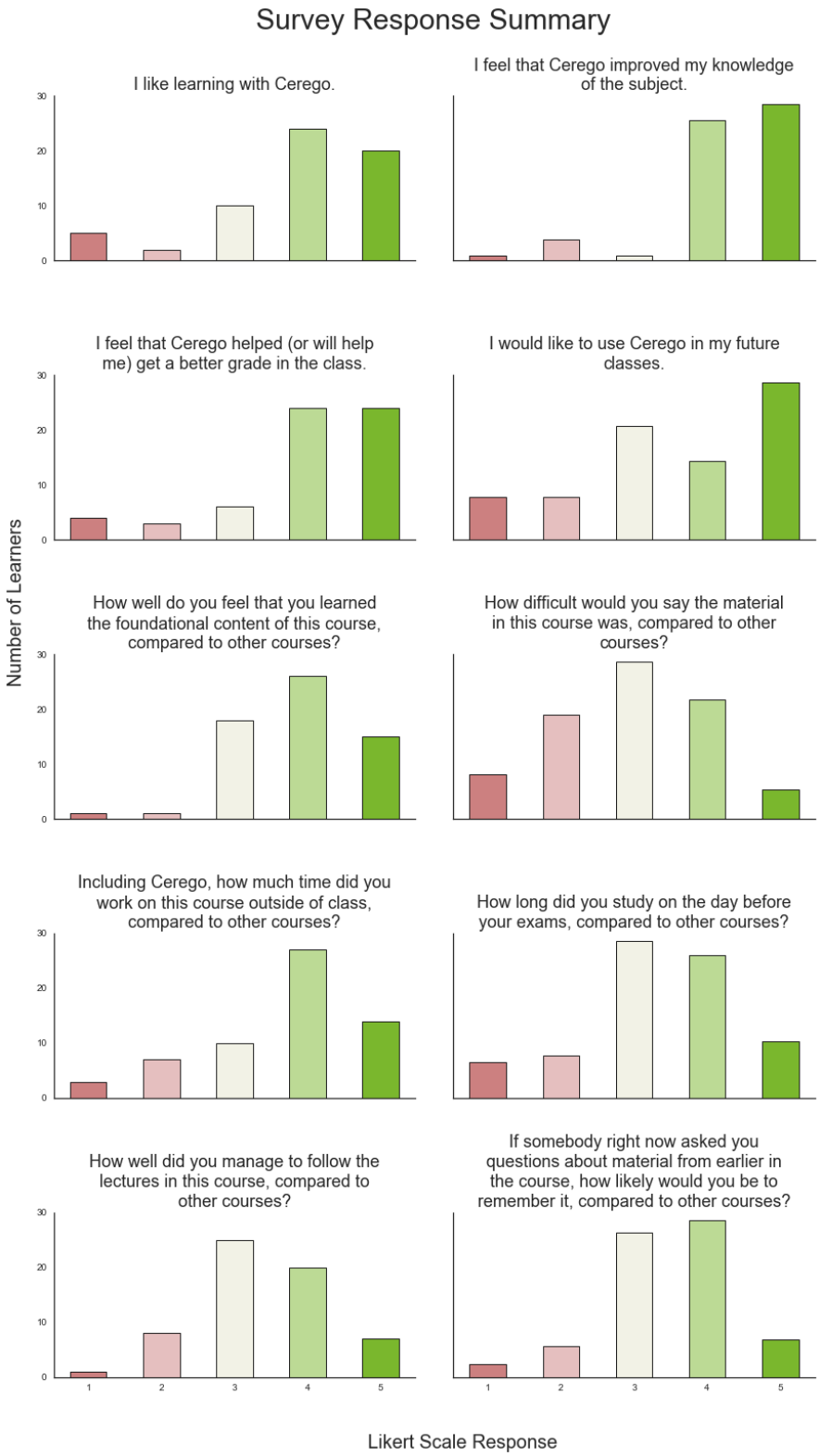

Figure 8: Visual summary of survey responses.

+0.41 response rating; p values starting at 0.043 and averaging 0.154) for higher agreement by Pell-eligible students.

In terms of the effects on studying, 44% v 15% agreed that they followed the lectures better in this course than other courses - but as discussed above, learners who had built more retention on Cerego were significantly more positive on this question. In two other questions designed to measure how Cerego affected the study schedule we found little evidence that Cerego caused an excessive workload (high Cerego users were no more likely to say they spent more time studying, even when Cerego was specifically included as part of that study time). In fact, we found some suggestion that Cerego might lead to a healthier distribution of study time, away from pre-exam 'cramming'. For all respondents, ratings of time spent cramming the night before the exam trended lower than ratings of overall study time outside class ($p = .072$). Most importantly, ratings of cramming trended higher for high GPA students ($p = .082$) but lower for those with high Cerego retention ($p = .159$).

Though inconclusive, this pattern is consistent with a scenario in which Cerego reduced the emphasis on cramming as a way to achieve high grades on the exam. The positive relationship between prior GPA and reported cramming implies that on average, cramming may be associated with higher exam scores in previous courses, but a direct test of the effects of reported cramming and Cerego retention on exam score for this course found the opposite. Multiple regression found Cerego to be a very strong predictor of exam score for survey respondents (+19 percentage points per level built; $p < .001$) but reported time spent cramming was weak to non-existent (+2 percentage points, $p = .200$). To the extent that grading should incentivize and reflect knowledge that is earned and lasting, rather than temporarily available for the day of the exam and quickly forgotten, this course did an excellent job. One conclusion is that adaptive learning approaches such as Cerego should be integrated more widely into college courses to further improve this.

5.4 Acknowledgements

We would like to thank the Bill & Melinda Gates Foundation for their generous funding of the Next Generation Courseware Challenge (NGCC), for which this study was performed. The NGCC seeks to improve access to, and measure the effectiveness of, adaptive learning systems in US colleges, and is especially focused on improving outcomes for underserved and at-risk populations such as minority, Pell-eligible or first-generation students. Cerego is proud to be a part of the NGCC, and deeply committed to scientifically testing and developing a platform that is genuinely effective for learners.

SRI and the Open Education Group provided vital support for ethics approval, data gathering and storage. We are also deeply grateful to the University of Hawai'i at Manoa for their efforts in combining and anonymizing the data for this study, and especially for their extra efforts in building such an unusually rich and informative dataset. In particular we would like to thank Professor Liang Wang who taught the Macroeconomics module in this study and integrated Cerego into his classes: his passionate commitment to his students is both evident and inspiring. Finally, we are indebted to the students; especially those who completed the optional survey component.

5.5 About the Author

Dr. Iain M. Harlow is VP of Science at Cerego, a data-driven learning company. He has a Ph.D. in neuroinformatics from the University of Edinburgh and has spent the past ten years as a researcher and data scientist studying - and working to improve - human memory and learning.