Cerego Insights

Measuring learning and potential

Iain M Harlow

September 26, 2018

Contents

1 Executive Summary

Cerego Insights provide learners and instructors with rich and meaningful measures of their state of knowledge, approach to learning, and ability to master new information. Cerego provide three primary metrics:

- *Knowledge* reflects what someone knows (breadth) and how well (depth).
- *Diligence* reflects the approach taken to learning.
- *Agility* reflects how quickly new information is mastered.

These metrics are designed to provide a richer, more useful, and more predictive set of insights into a learner than can be provided by a test score or a personality quiz. In this white paper we outline in detail what each metric is designed to reflect, how it is measured, and why it is valuable.

Knowledge

Knowledge in Cerego measures both the *breadth* and *depth* of a learner's knowledge: it reflects what somebody knows, and how well they know it. Depth of knowledge (retention) is especially important to higher-order thinking and understanding, and difficult to measure on a test. Depth of knowledge is strongly and consistently predictive of external outcomes such as exam scores and post-tests (Figure 1), including measures of a learner's deeper understanding.

Diligence

Diligence measures a user's *approach* to learning. Diligent learners tend to complete tasks and goals, log in and learn more consistently and in shorter sessions, and engage thoughtfully with the to-be-learned material, especially when it is challenging. All of these behaviors lead to better *future* performance on the same material, i.e. they are strategies that correspond to effective learning.

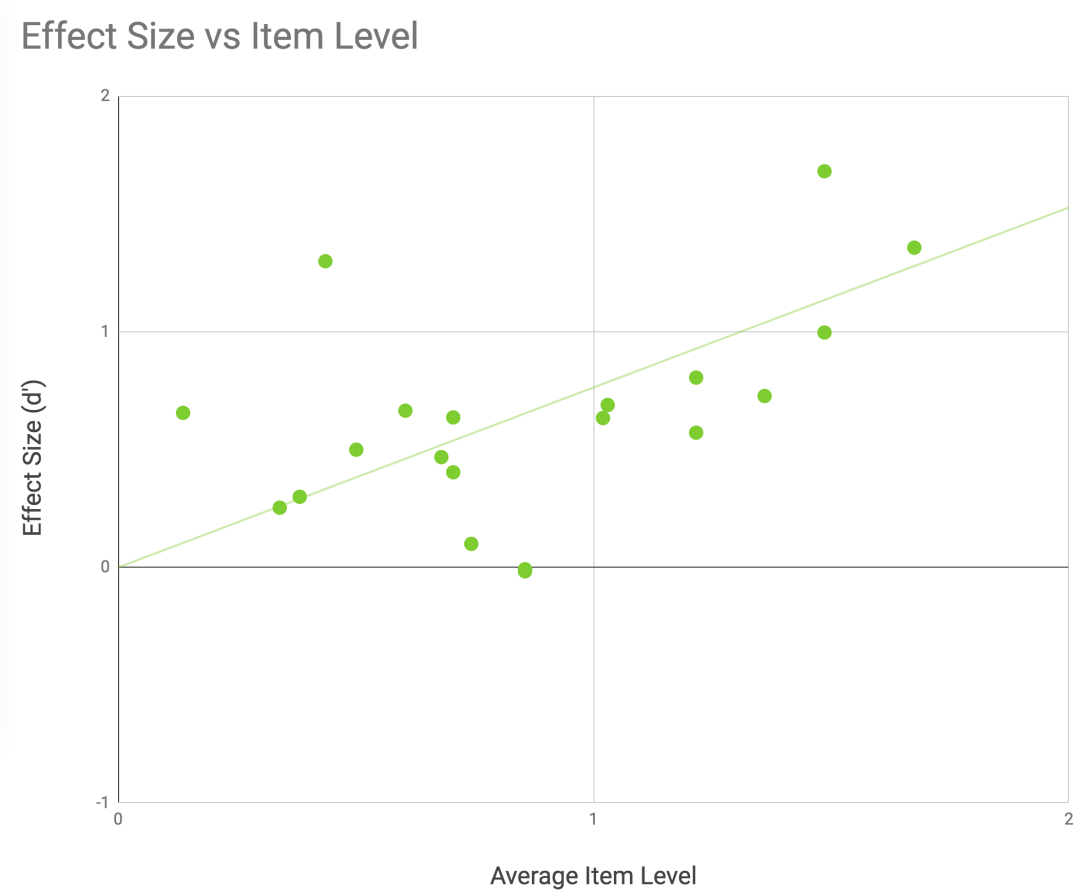

Figure 1: The relationship between measured real-world effect size (e.g. improvement on an exam) and average item level (depth of knowledge) in Cerego. Points represent a comparison from an efficacy study. Knowledge predicts real world outcomes strongly, with each level of retention corresponding to +0.76 standard deviation units of improvement.

Agility

Agility reflects the *potential* of a learner, and measures how easily they learn and retain information. Agility is content-agnostic: whereas knowledge varies across material (e.g. one may be an expert in aerodynamics but not psychology), agility reflects the general learning potential of the individual. Agility is conceptually and statistically validated across hundreds of millions of interactions in Cerego, and predicts real-world outcomes such as training retention and exam score.

2 Measuring learning

Cerego's Insights do not exist in an epistemic wilderness. For centuries, individuals have been scored, graded and awarded qualifications primarily through performance on exams and tests of various kinds. Testing has some advantages - it is a convenient and familiar approach to measuring learning, and it does not generally require technology - but it also has a number of downsides.

One downside is that testing provides only a moment-in-time snapshot of what somebody knows. Human memory for new information decays rapidly; studies have shown for example that anywhere from 73% to 86% of newly learned written or spoken material may be lost within a few weeks [1, 2]. As a result, testing encourages cramming: short-term memorization of material for the purposes of passing a test. Cramming is a notoriously ineffective way of building lasting retention [3, 4]. Worse, a test score will not distinguish a learner who has built lasting retention from one who has crammed for a test and forgets the material shortly afterwards. Test scores tell us what somebody *knew* but they cannot tell us what somebody *knows now or will know in the future.*

Incentivizing cramming is not the only downside to testing. Moment-in-time tests such as exams disadvantage some students, especially those 10-25% with test anxiety [5] or learning difference factors such as dyslexia. Exams and tests (and resulting measures of ability, such as grade point averages, GPA) also correlate with economic and demographic variables. For example, SAT results correlate strongly with minority status and family income [6]. Large scale studies across thousands of students [7] find ACT or SAT submission to be a poor predictor of college success, compared to high school GPA which measured achievement over time rather than on a single test.

In short, our current approach to testing is an imperfect measure of ability, knowledge and potential.

3 Knowledge

In Cerego, knowledge is not measured through test scores, but by the amount of *retention* a learner has built up. Retention refers to the length of time a memory will last before it has faded too much to use; memories in Cerego with higher retention correspond to knowledge that will last, and be accessible to the learner in the future. Most newly learned material begins at low retention, and is therefore quickly forgotten, but retention for material is increased by reviewing the material periodically - prompted by Cerego's scheduling algorithm. This algorithm is based on myriad factors, but notably *distributed learning* and *retrieval practice*, two of the most effective ways to improve long-term retention [4].

3.1 Measuring Knowledge

Knowledge retention in Cerego is logarithmically transformed and reported as a *level* for every concept a user learns, on a scale of 0-5 (non-integer values, such as 1.7, are allowed). The position on the scale corresponds to the slope of the memory's decay curve: memories with higher retention levels follow a shallower slope, and can be recalled more easily, more automatically, and for longer. For example, memories built to level 1 will last for weeks, while memories built to level 3 or more will last for months or years.

Attaining a higher retention level for a concept does more than simply extend the length of time it is available. Increasing the length of time a learner can go between reviews reduces the amount of time they must spend maintaining that knowledge, allowing them to actively maintain a *broader* range of knowledge at once. Simultaneously, increasing retention level *deepens* knowledge, growing the kind of familiarity with and automaticity for concepts required for true expertise.

Experts in a subject are largely distinguished by the depth and breadth of their domain knowledge. In fact, the value of their expertise - superior understanding, analytical and creative abilities - arises directly *because* of this knowledge [8]: experts see systems differently to non-experts precisely because their knowledge leads them to recognize more sophisticated patterns and relationships [9, 10]. Knowledge is also reinforcing: absorbing new information becomes easier when you have an existing framework of knowledge to understand and interpret it [11, 12].

The purpose of Cerego is to help learners build expertise (rather than simply pass

a test). The knowledge metric in Cerego is constructed from a sum of retention levels across all the content being learned; it therefore captures both *breadth* and *depth* of knowledge in a single summary statistic.

3.2 Validating Knowledge

A key validation for Cerego's knowledge metric is the fact that it predicts performance on outside measures such as exams. Consistent with the research showing that knowledge is critical for understanding and expertise, these performance improvements are not limited to factual multiple-choice question, but are present and even greater when understanding and analytical skills are directly tested.

The Next Generation Courseware Challenge, funded by the Bill & Melinda Gates Foundation, provided an opportunity to validate Cerego's knowledge metric in a real-world context. Students at the University of Hawai'i at Manoa used Cerego as part of a Macroeconomics college course [13]. Knowledge built in Cerego for these learners corresponded to improved exam scores (Figure 2), and this improvement grew over time as Cerego learners found it easier to follow the lectures as new material was introduced.

Even after accounting for student abbility using prior GPA, each level of retention built in Cerego was associated with a 12% improvement on the final exam, and a larger 16% improvement on the more challenging analytical questions that test higher order thinking and understanding.

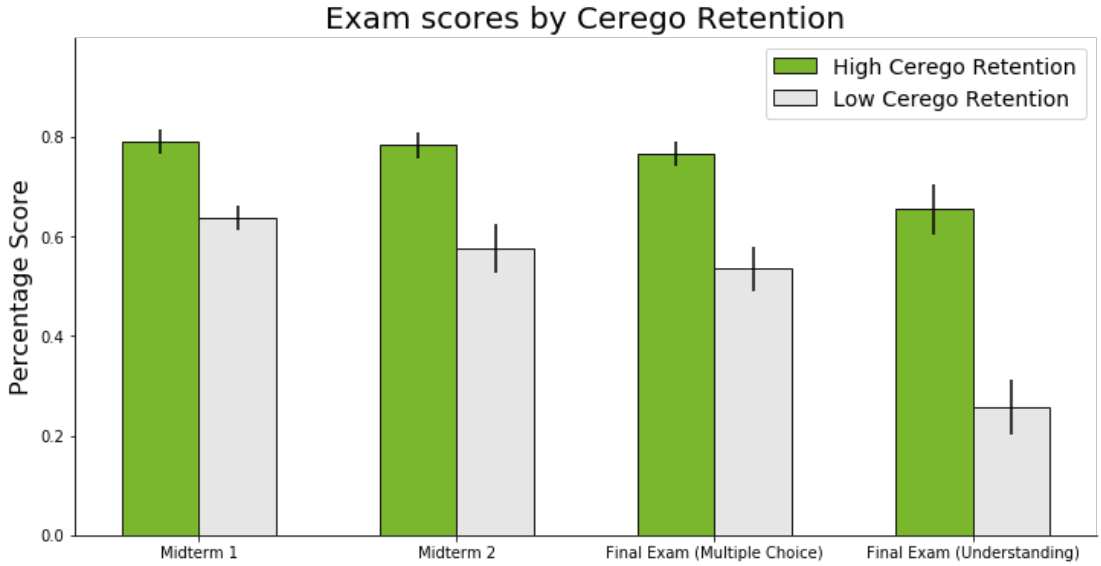

Figure 2: Scores on two midterm exams and one final exam at UH Manoa revealed a substantial and increasing gap in educational attainment between students who built high (top third) and low (bottom third) Cerego retention. The visible trends of greater exam performance by students who built high Cerego 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.

Other studies have found a similarly reliable link between the knowledge measured by Cerego and subsequent real world outcomes, for a variety of material and ages. For example, similar evidence of knowledge transfer has been observed for high school students in Florida [14], dental school students at New York University [15], and Syrian refugee childreen in Turkey [16].

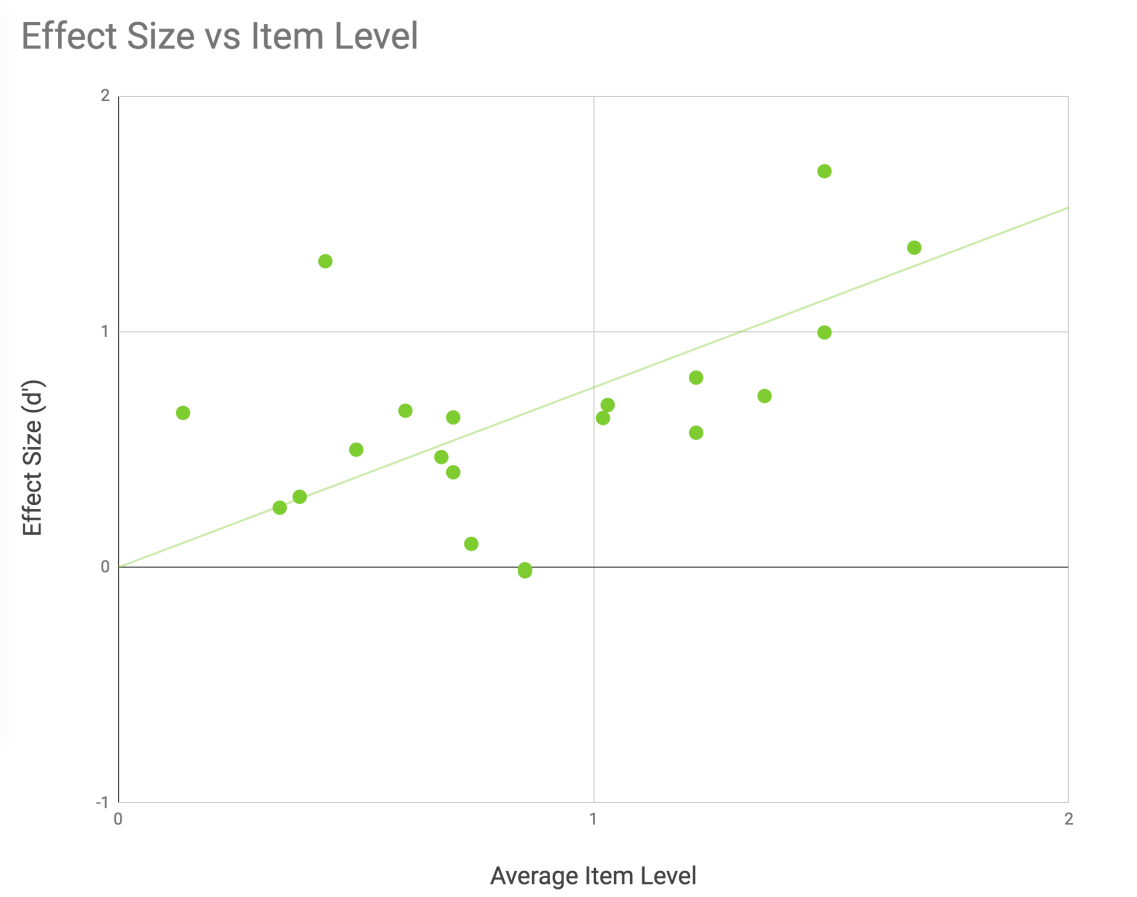

Figure 3: The relationship between measured real-world effect size (e.g. improvement on an exam) and average retention level achieved in Cerego. This figure is illustrative: points each represent a comparison from a pilot, peer-reviewed study, case study or internal research but vary in the comparison type (e.g. dosage effects vs randomized controlled trials) and may include some non-independence (e.g. two points may share a control group). Overall, Cerego's knowledge retention tends to predict real world outcomes strongly, with outcomes improving for 19 of 21 comparisons, and each level of retention corresponding to around +0.76 standard deviation units of improvement in a real world outcome.

More broadly, across case studies, pilots and peer-reviewed research, the effect size measured - the size of the benefit to using Cerego - is closely related to the average retention level built in the system. On average, external outcomes (variously exam results, post tests, pass rates or other real-world measures) rise by around 0.76

standard deviations per level of retention gained in Cerego (Figure 3).

Knowledge metrics in Cerego capture a learner's current breadth and depth of expertise, not simply performance on a past test. They show strong and consistent relationships with external measures of ability, especially those that measure understanding.

4 Diligence

Consistently, research finds conscientiousness to be one of the most robust predictors of both academic and professional success [17, 18, 19, 20]. Cerego's diligence measure is designed to capture aspects of conscientious learning behavior that correspond to effective learning strategy. Broadly speaking, higher diligence indicates that a learner is more consistently, deliberately and effectively using their time when learning and reviewing in Cerego, with correspondingly improved learning outcomes.

4.1 Measuring Diligence

The factors that go into Cerego's diligence metric are empirically derived: they reflect how closely a learner follows learning strategies that are proven to be effective. For example, "little and often is the key": Retention is most effectively built when learners review content frequently in short sessions of a few minutes. Reviews are also more effective when they are sufficiently challenging, landing in a window known as *desirable difficulty* [21]. Cerego's scheduling algorithms identify this window for each concept and schedule reviews for each learner accordingly. The diligence metric captures how regularly each user reviews the content they are learning, and how closely they follow their optimized, custom schedule.

Diligence also captures other deliberate behaviors that improve learning. For example, engaging deeply with a question before attempting to answer it is more effective for building retention than simply recognizing the correct multiple choice response [22]. Cerego's learning platform encourages this engagement by separating the question from the answer choices on many review quizzes, encouraging learners to interrogate their memory and consider possibilities *before* selecting a response. Similarly important moments for engagement exist at other points in the learning process - for example pausing to understand the correct answer after making a mistake - these are also tracked and included in the diligence metric.

The precise mix of factors in the dilgence metric will update over time, based on how effective the underlying behaviors actually prove to be. In general, a learner who responds to notifications, schedules a few minutes each day to review, and carefully considers the content and quizzes, will tend to have a high diligence rating. A learner who delays learning until shortly before a deadline and skips rapidly through the content will tend to have a lower diligence rating.

4.2 Validating Diligence

The factors that underlie the diligence metric are derived from learning and memory science research, and are quantified and measured in the Cerego system. Following the Cerego schedule more closely is one straightforward and important factor, as would be expected since the schedule optimizes review times for desirable difficulty and effective learning. Adherence to the schedule alone accounts for accuracy differences on future quizzes of more than 12 percentage points.

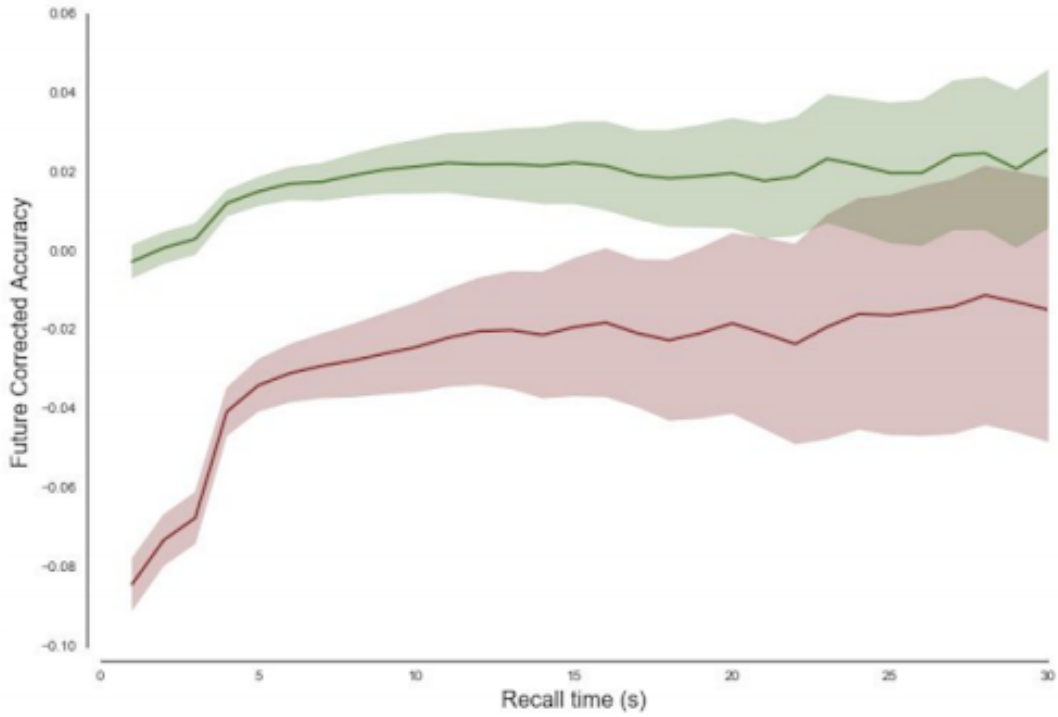

Figure 4: The relationship between engagement, i.e. time spent attempting to recall a concept, and *future* accuracy for that concept. Future accuracy improves when learners spend 5-10 seconds considering the question before answering, whether the current quiz response ends up being correct (green upper line) or not (red lower line). Future accuracy is corrected for factors such as the time elapsed since it was last seen.

Another important component of the diligence metric, the amount of time spent attempting to recall the answer, is also strongly beneficial for learning in Cerego.

Figure 4 shows that spending 5-10 seconds considering a question pays off with 2-6 percentage point improvements in *future* accuracy. This is true whether or not the response on the current quiz ends up being correct or not.

This kind of deliberate, effective engagement with the content in Cerego varies within learners, too - sometimes we approach a task thoughtfully and other times not. Generally speaking, learners consider each question for longer (and therefore more effectively built lasting retention) earlier in a session, when they typically have more energy and motivation to learn (Figure 5).

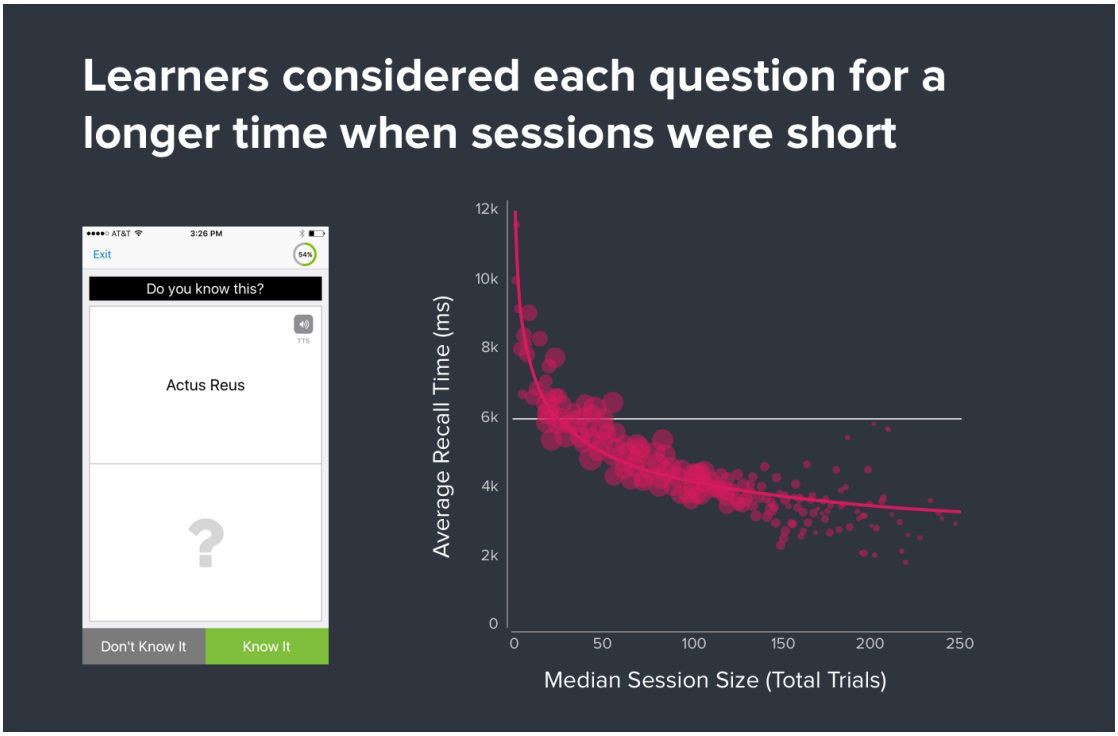

Figure 5: Time spent considering each question - an important predictor of learning - varies consistently with the length of the learning session. Spending 5-10 seconds on each question improves learning (Figure 4), but learners tend to skip through the quizzes faster than this if they review more than 20-30 concepts at a time.

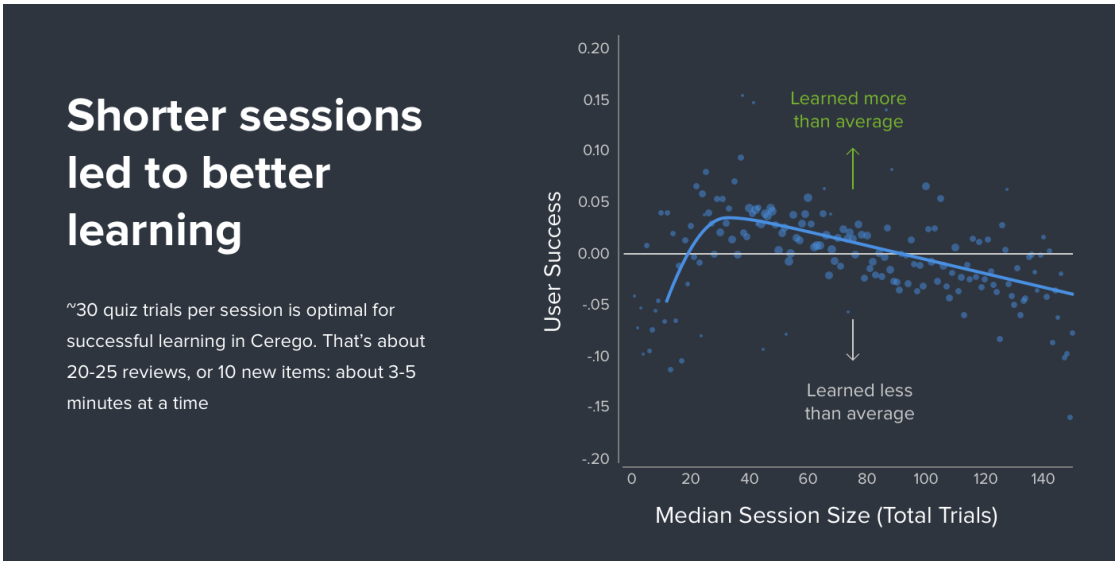

Figure 6: Learners' progress towards lasting retention was considerably faster when they averaged shorter review sessions. Learners with a median session length of 3-5 minutes see future accuracy improvements of 5-10 percentage points compared to those who spend longer studying without breaks.

Over longer sessions motivation and alertness deplete and learners engage less and less effectively. In Cerego, learners who spend around 3-5 minutes learning before taking a break - i.e. those who review "little and often" - see much more efficient learning as a result (Figure 6). Once again measuring success as average *future* accuracy for the material, these "little and often" learners saw roughly 10 percentage point improvements compared to those who studied for blocks of 30 minutes.

5 Agility

The learning agility metric in Cerego is, at its core, an accuracy measure. Just as an exam score generally reflects an average accuracy across all the questions on an exam, the agility score reflects an accuracy across all the review interactions performed in Cerego. The agility score differs from a simple test score in two important ways, however, both of which improve its reliability and value:

- Agility is calculated across a much greater number of interactions than a typical exam score. The average number of reviews used to calculate a learner's agility is 450. This mitigates the effects of random factors such as lucky guessing or motor error ("fat finger" responses). Even more importantly, this may also mitigate systemic sources of error such as exam anxiety or concentration factors, since the learner has control over when they review, and each individual review has lower stakes than in a shorter exam.

- Agility takes into account key influences on memory such as the difficulty of the concepts being learned, the types of quizzes faced, and the length of time since a concept was last seen. This further mitigates factors such as question difficulty (which can be a systemic factor in exam scores, e.g. between years or exam versions) and de-emphasizes cramming.

It is important to emphasize that learning agility is *not* an IQ measure, or a rating of the general intelligence or ability of a user. It is a measure of the user's ability to master new information: how fast they learn, and how well they tend to retain information.

5.1 Measuring Agility

Learning agility is a parameter that comes directly from Cerego's predictive memory tracking AI. Before a learner completes a review of a concept in Cerego, the algorithms at the core of Cerego's memory tracking AI generate an estimate of the chance they will answer successfully:

$$P(correct) = f(\vec{x})$$

Function $f(\vec{x})$ is a scientifically-constrained model of memory decay and stabilization, updated by an online machine learning algorithm, in which memory recall probability decays along a power decay curve over time between reviews. This power decay (as opposed to an exponential decay) is observed both in recent memory research [23, 24, 25] and (very clearly) in Cerego's own empirical data of millions of memories. The input vector $\vec{x}$ contains the context and learning history required to calculate the recall probability, including:

- The *learning history* for this user and this concept, such as the number of reviews completed, reaction times, estimates of retention, timestamps of previous interactions and other useful information.

- The difficulty of the specific *concept* being quizzed, learned across all interactions any user has had with that concept.

- The difficulty of the *quiz type* being attempted, learned from all previous interactions any user has had with that quiz type. For example fill-in-the blank questions are usually more difficult than multiple choice quizzes, even if they are testing the same concept.

- The *agility* of the user, learned across all interactions the user has had with any content. Higher agility learners are more likely to answer correctly, across content, quizzes and circumstances.

- The *date and time* of interest, often the current moment but can be changed to predict the outcome on any given date.

After the model generates a prediction, the user attempts the review and either succeeds or fails (or in some cases partially succeeds). Based on the outcome the online predictor will immediately update not only the learning history for that memory, but also its estimates of quiz and concept difficulty, and user agility. Hence each single review in Cerego updates the model's understanding of how difficult the content is and how users differ in learning agility, enabling it to produce more accurate predictions for future reviews.

The agility variable, measured for each learner and updated after every learning interaction, provides a concise and valuable measure of each user's learning rate. Where diligence measures effort, or the *approach* to learning, which might change week to week, agility can be thought of as a measure of learning *potential*.

5.2 Validating Agility

Cerego have made learning agility available to users of the Cerego system following careful validation of the metric. This longer section explains why learning agility is a valid and useful concept: It is statistically necessary to explain memory data; it is conceptually distinct from knowledge (including prior knowledge built outside of Cerego's system), and it is empirically consistent and related to measures of learner success outside Cerego.

5.2.1 Statistical validation of learning agility

Does the agility metric capture something consistent and real about a learner's own ability? Or does it simply capture noise that over time averages out? To answer this question, we can compare two models: A simple version in which all learners have the same agility metric, and a more complex one (as detailed above) that includes a separate agility metric for each learner.

The models are compared on predictions they make for future, i.e, unseen, reviews. If the simpler model is more correct, and agility is capturing primarily random non-repeatable factors such as lucky guesses or extra non-Cerego reviews, it should predict the accuracy of future reviews better than the more complex model, which will a) tend to overestimate performance for users who have been historically more "lucky" and b) have fewer degrees of freedom to accurately measure real factors such as the difficulty of items.

The opposite pattern occurs, however (Figure 7): The model performs vastly more accurately when it includes the previously-estimated user agility values. This provides overwhelmingly strong evidence that user agility metrics vary consistently and robustly across users, predict future performance for each user, and do not reflect random differences or past performance for specific items.

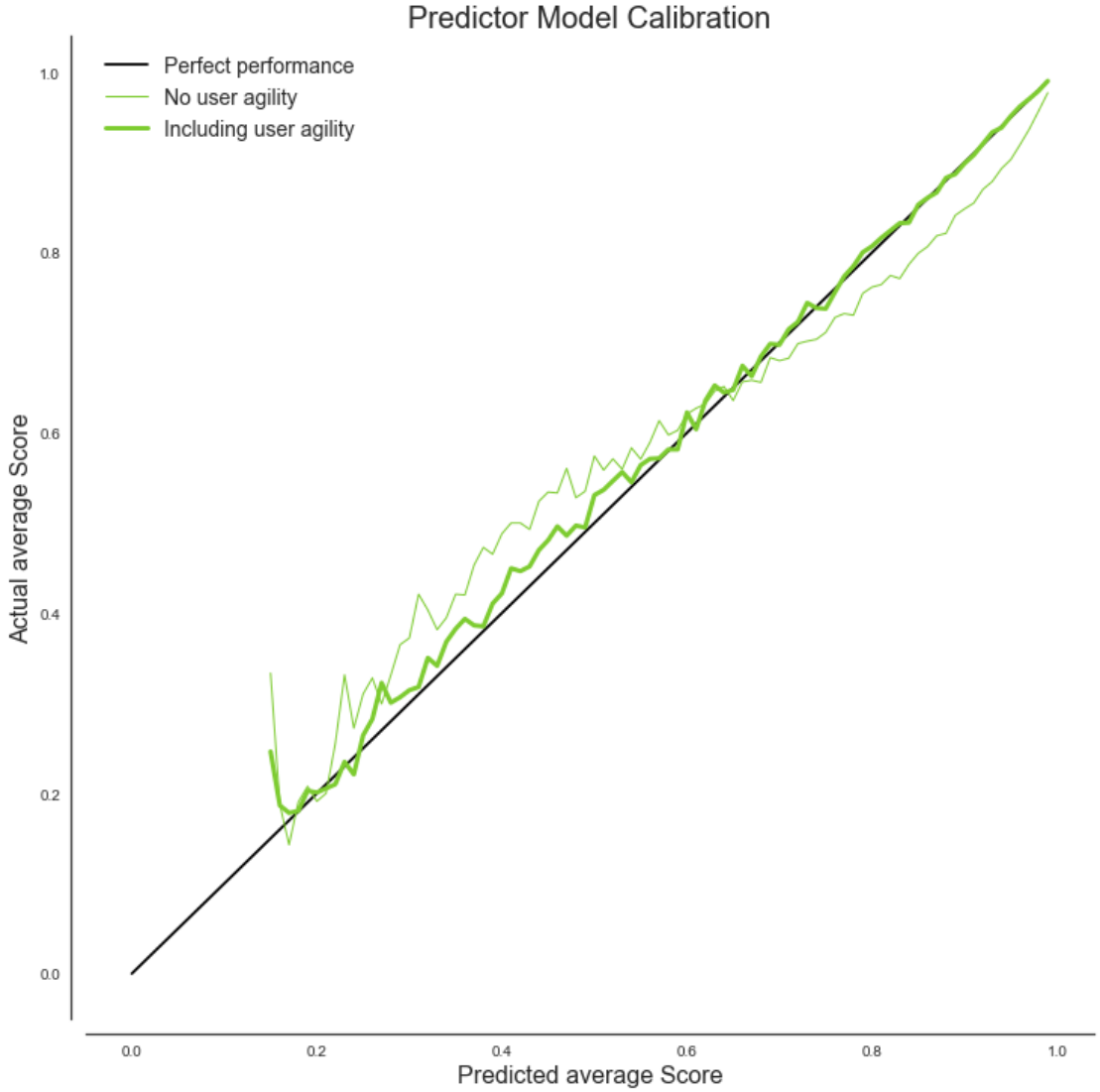

Figure 7: The predictor model is vastly more accurate at predicting new unseen reviews when it includes the user agility metric than when it assumes learners are similar in ability. Log-likelihood for 407095 reviews over a 4-day period in July 2018 improves from -192847 to -162509 with the inclusion of 5032 previously estimated user agility variables [Chi-squared test: $D = 60675, df = 5032, p << 0.001$]. In fact, as is apparent from the figure above, the agility-including model is close to perfect for reviews with a greater than 50% chance of being correctly answered (which is to say, more than 90% of all reviews). The data provide overwhelming evidence that users vary in learning agility, and that this variation is consistent (predictive of future performance in the system).

5.2.2 Conceptual validation of learning agility

Since agility is fundamentally a measure of accuracy, it is intuitively reasonable to imagine that prior knowledge for some material should improve a user's measured agility (by way of improving their accuracy on questions relative to what our system - initially naive to their prior knowledge - would expect).

Prior knowledge should help most on questions for which the learner has seen the answer but others have not. This is because of a mismatch between the learner's true retention (high) and Cerego's estimate (unknown, initially assumed to be low). This combination is most likely to be true when:

- Learners have reviewed at most once or twice, i.e. when Cerego has little evidence of their real retention and
- There is a high overlap between the questions asked in Cerego and the concepts previously encountered by the learner.

In practice, the predictor engine does three things that should quickly mitigate any agility-inflation caused by prior knowledge.

First, agility is updated based on performance-relative-to-expectation - where expectation is in turn based on past performance, including agility. This means that if agility has been overestimated from the first few reviews, the system will expect unrealistically high accuracy later on. But unlike the first review, now the model is also less naive about the learner's retention for the material: it recognizes that the learner is likely to have high retention for the material and the gap between the reality (learner has built high retention outside of Cerego) and Cerego's estimate (retention estimates based on performance in Cerego) has started to narrow. Since the model has overestimated the learner's agility, but started to catch up on its estimate of their retention, the model's expectations for performance beyond the first few reviews become increasingly hard for the learner to match in reality. The result is that the agility estimate will tend to decrease back towards its true underlying value.

Second, the same effect happens on any reviews for content that is a less close match to the learner's existing knowledge. Even within a set, some items/quizzes will usually be at least somewhat new to a learner with prior knowledge. A similar dynamic will occur as for later reviews - the model will expect the learner to overperform on these questions (relative to their peers) just as much as they

overperformed on the quizzes they had explicit prior knowledge for, so that the actual performance shifts agility estimates back down towards their true value. Notably, this does more than simply 'average out' the learner's agility estimate, since the less accurately answered quizzes will be seen more often due to their lower retention.

Finally, the model is in the first place quite conservative about estimating agility from a small number of reviews. Agility estimates are strongly anchored towards the average, and it takes a substantial amount of evidence (in the form of reviews consistently answered more accurately than expected) to move the estimates. This means that the circumstances when agility estimates would be most affected (the first review or two for each item, and when learners are only studying content they are already familiar with) are exactly the circumstance (few reviews) when agility is updated most conservatively, further dampening any errors.

All of this is to say that the model is designed to adapt to prior knowledge, and to estimate agility across a wide range of content and time. Fundamentally, the model is designed to independently estimate

- a learner's current knowledge retention for specific content, and
- their learning agility across all content.

A learner with prior knowledge enters the system with greater knowledge retention for specific content, and so they will show above-average performance on that content (relative to peers) but not in general across all content. Thus, over time and content the model will tend to learn this true pattern: higher knowledge retention for those specific items, not generally higher learning agility.

5.2.3 Empirical validation of learning agility

The reasoning above suggests that prior knowledge could affect agility estimates to some degree, and under particular circumstances: the effects should be small and decrease as the number of reviews increases. Empirical data from the Cerego system provides some additional evidence around the impact of prior knowledge. Firstly, there is no evidence that learning occurs outside of Cerego to a sufficient extent to significantly affect performance within the system. If this did happen - learners regularly reviewed material outside of Cerego, improving their

performance when later quizzed within the system - this should manifest as an asymmetry in the model calibration plot (Figure 7). Specifically:

- Reviews of concepts that have not been recently seen within Cerego should be most affected, since these longer gaps outside the system are when non-Cerego review is most likely to have taken place;

- These reviews are disproportionately in the lower-predicted-accuracy range of the curve, since the time elapsed since learning in Cerego means predicted performance will be low;

- This effect should not be balanced out by underperforming reviews since the *shape* of the memory decay curve is known to be a scientifically-derived power decay curve and there should not plausibly exist reviews for which outside behavior has significantly *reduced* memory;

- The result is that we should see an overperformance in the model calibration curve (i.e. the curve should be shifted above the perfect x=y line) specifically for reviews with lower expected accuracy, i.e. the left part of the curve.

Such a pattern is not in fact especially evident in the data. There is some small hint at the very extreme left (the least well remembered 5% of reviews are in reality remembered slightly more accurately than the model expects), which might reflect outside study of material not seen in Cerego for a long time. It may also, however, reflect a limitation of the model at extremely low activations, where rounding errors begin to have a larger effect; or a performance floor introduced by non-mnemonic factors such as process-of-elimination responses to multiple choice questions; or some combination of these factors.

In practice, therefore, reviewing outside of Cerego appears to occur too infrequently or ineffectively to have any meaningful effect on the predictive model (and, by extension, the agility parameter of that model).

It is also possible to directly compare the agility of learners who are new to material, and learners who should be more familiar with the same content. The African Leadership University use Cerego during the application process for new students. New applicants engage with a Cerego assignment for a few days (level 1.5), providing the University with additional insight into their potential that may otherwise be obscured by differences in national, educational or economic background.

Before launching the application assignment, ALU tested it with a sample of existing ALU students. Crucially, the content being learned by both applicants and existing students was entirely about ALU itself - their mission, structure, approach to learning and finance. Existing ALU students should presumably be more familiar with the content than brand new applicants. Agility scores for the two groups however were not significantly different [$t(274) = 1.058$; $p = .291$]. Furthermore, agility for the existing students was not significantly related to their performance on quiz-first, Question-and-Answer items that are most likely to reflect prior knowledge. Instead, their agility was significantly related to their overall ALU GPA: students with higher grades at ALU tended to also show higher agility scores within Cerego.

6 Summary

Most widely used measures of a person's ability - exams, interviews, personality quizzes - measure performance at a single moment in time. Measuring ability in this way is convenient, but introduces hidden bias and large elements of chance, and limit the ability to make *forward-looking* judgments of how a person will perform a task in future.

Cerego Insights are drawn from predictive, scientific models of human behavior and ability, over an extended history of interactions. Each of the metrics, their underlying components, and the predictions they make, are validated against external outcomes and a database of billions of interactions. Each is subject to continued testing, refinement, and improvement, and additional insights will be introduced as they reach and exceed similar standards of reliability and validity.

Cerego Insights are designed to be predictive, valid, unbiased and scientific - and form an integral part of our mission to measure and improve human potential.

About the Author

Dr. Iain M. Harlow is VP of Science at Cerego. 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.

References

[1] Francis J. Di Vesta and Deborah A. Smith. "The pausing principle: Increasing the efficiency of memory for ongoing events". In: *Contemporary Educational Psychology* 4.3 (1979), pp. 288–296. ISSN: 0361-476X. DOI: https://doi.org/10.1016/0361-476X(79)90048-1. URL: http://www.sciencedirect.com/science/article/pii/0361476X79900481.

[2] Damon Krug, T. Brandon Davis, and John A. Glover. "Massed versus distributed repeated reading: A case of forgetting helping recall?" In: *Journal of Educational Psychology* 82.2 (1990), pp. 366–371. DOI: 10.1037/0022-0663.82.2.366.

[3] Doug Rohrer and Harold Pashler. "Increasing Retention Without Increasing Study Time". In: *Current Directions in Psychological Science* 16.4 (2007), pp. 183–186. DOI: 10.1111/j.1467-8721.2007.00500.x. eprint: https://doi.org/10.1111/j.1467-8721.2007.00500.x. URL: https://doi.org/10.1111/j.1467-8721.2007.00500.x.

[4] John Dunlosky et al. "Improving Students Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology". In: *Psychological Science in the Public Interest* 14.1 (2013), pp. 4–58. DOI: 10.1177/1529100612453266.

[5] Angus S. McDonald. "The Prevalence and Effects of Test Anxiety in School Children". In: *Educational Psychology* 21.1 (2001), pp. 89–101. DOI: 10.1080/01443410020019867. eprint: https://doi.org/10.1080/01443410020019867. URL: https://doi.org/10.1080/01443410020019867.

[6] Saul Geiser. *The growing correlation between race and SAT scores: New findings from California*. Berkeley, CA, USA, 2015. URL: https://cshe.berkeley.edu/sites/default/files/publications/rops.cshe_.10.15.geiser.racesat.10.26.2015.pdf.

[7] William C. Hiss and Valerie W. Franks. *Defining promise: Optional standardized testing policies in American college and university admissions*. 2014. URL: https://offices.depaul.edu/enrollment-management-marketing/test-optional/Documents/HISSDefiningPromise.pdf.

[8] R. Taconis, M.G.M. Ferguson-Hessler, and H. Broekkamp. "Teaching science problem solving: An overview of experimental work". In: *Journal of Research in Science Teaching* 38.4 (2001), pp. 442–468. DOI: 10.1002/tea.1013. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.1013. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.1013.

[9] Adriaan D. De Groot. *Thought and choice in chess*. The Hague, The Netherlands: Mouton Publishers, 1965. ISBN: 90-279-7914-6.

[10] Michelene T. H. Chi, Robert Glaser, and E. Rees. "Expertise in problem solving". In: *R.J. Sternberg (Ed.), Advances in the psychology of human intelligence* 1 (1982), pp. 7–35.

[11] Margaret E. Beier and Phillip L. Ackerman. "Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: promising results in a real-world learning environment". In: *Psychology and Aging* 20.2 (2005), pp. 341–355. DOI: 10.1037/0882-7974.20.2.341.

[12] John R. Anderson. "Effects of prior knowledge on memory for new information". In: *Memory & Cognition* 9.3 (1981), pp. 237–246. ISSN: 1532-5946. DOI: 10.3758/BF03196958. URL: https://doi.org/10.3758/BF03196958.

[13] Iain M. Harlow. *Beyond the Foundations: A quantitative investigation of Cerego's impact on knowledge transfer and understanding*. San Francisco, CA, USA, 2018. URL: https://www.cerego.com/hubfs/Blog%20Media%20and%20Resources/PDF/Beyond%20the%20Foundations.pdf.

[14] Bruce D. Homer and Jan L. Plass. *Summary of Cerego Academica pre-pilot study*. New York, NY, USA, 2013. URL: https://cdn2.hubspot.net/hubfs/2480790/Blog%20Media%20and%20Resources/PDF/K12Outcomes.pdf.

[15] Johanna Warshaw et al. "Mastering anatomy: Using Cerego as a teaching tool". In: San Diego, CA, USA: Proceedings of the American Association of Anatomists, 2018.

[16] Selcuk Sirin et al. "Digital game-based education for Syrian refugee children: Project Hope". In: *Vulnerable Children and Youth Studies* 13.1 (2018), pp. 7–18. DOI: 10.1080/17450128.2017.1412551. eprint: https://doi.org/10.1080/17450128.2017.1412551. URL: https://doi.org/10.1080/17450128.2017.1412551.

[17] Murray R. Barrick and Michael K. Mount. "The Big Five personality dimensions and job performance: A meta-analysis". In: *Personnel Psychology* 44.1 (1991), pp. 1–26. DOI: https://dx.doi.org/10.1111/j.1744-6570.1991.tb00688.x.

[18] Adrian Furnham, Tomas Chamorro-Premuzic, and Fiona McDougall. "Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance". In: *Learning and Individual Differences* 14.1 (2003), pp. 47–64. ISSN: 1041-6080. DOI: https://doi.org/10.1016/j.lindif.2003.08.002. URL: http://www.sciencedirect.com/science/article/pii/S1041608003000359.

[19] Nicole Conrad and Marc W. Patry. "Conscientiousness and academic performance: A mediational analysis". In: *International Journal for the Scholarship of Teaching and Learning* 6.1 (2012). DOI: https://doi.org/10.20429/ijsotl.2012.060108.

[20] Sevag K. Kertechian. "Conscientiousness as a key to success for academic achievement among French university students enrolled in management studies". In: *The International Journal of Management Education* 16.2 (2018), pp. 154–165. ISSN: 1472-8117. DOI: https://doi.org/10.1016/j.ijme.2018.02.003. URL: http://www.sciencedirect.com/science/article/pii/S147281171730407X.

[21] Robert A. Bjork. "Memory and metamemory considerations in the training of human beings". In: *J. Metcalfe and A. Shimamura (Eds.), Metacognition: Knowing about knowing*. Cambridge, MA, USA: MIT Press, 1994, pp. 185–205.

[22] Giulia Galli. "What Makes Deeply Encoded Items Memorable? Insights into the Levels of Processing Framework from Neuroimaging and Neuromodulation". In: *Frontiers in Psychiatry* 5 (2014), p. 61. DOI: 10.3389/fpsyt.2014.00061.

[23] John T. Wixted and E. B. Ebbesen. "Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions". In: *Memory and Cognition* 25 (1997), pp. 731–739.

[24] Michael J. Kahana and Mark Adler. *Note on the power law of forgetting*. Philadelphia, PA, USA, 2002.

[25] Chris Donkin and Robert M. Nosofsky. "A power law model of psychological memory strength in short- and long-term recognition". In: *Psychological Science* 23.6 (2012), pp. 625–634. DOI: 10.1177/0956797611430961.