Parameter Correlations in the Predictive Performance Equation: Implications and Solutions

Michael Collins[1,2,3]
Florian Sense[3]
Michael Krusark[5]
Joshua Fletcher[6]
Tiffany Jastrzembski[1]
[1]Air Force Research Laboratory [2]ORISE [3]Wright State University [4]Infinite Tactics [5]L3Harris Technologies [6]Ball Aerospace

Abstract

Research of mathematical models of learning and retention have focused on accounting for an individual's performance across a variety of learning schedules (i.e., spaced and massed). The attempted goal of such research is to develop a model which can adequately predict human performance across a range of learning scenarios. However, little attention of this model development has focused on the interpretation of a model's best fitting parameters given the structure of a model's equations and its predicted performance values. The effect of this can lead to the development of models where the parameter values are correlated hindering a theoretical interpretation of performance. Here we examine the structure of the Predictive Performance Equation (PPE) and highlight portions of PPE's equations that lead to correlations across its free parameters. We propose a fix for these issues (Modified PPE) and conduct a formal model comparison showing the Modified PPE is simpler, has less parameter correlation and its best fitting parameters map on to identifiable aspects of an individual's performance.

Keywords: memory, learning, decay, spacing effect, mathematical modeling, model comparison, model identifiability

Introduction

Mathematical models of learning and retention are quantitative formulations of verbal psychological theories which attempt to account for and/or predict empirical data. One value of these mathematical formulations is the fact that all assumptions of a model are made explicit allowing for formal statistical evaluation. Furthermore, these mathematical models lend themselves to real-world applications, such as adaptive learning systems. Although the quantitative formulations of models have many benefits, care must be taken to ensure how these models are constructed, to ensure that a model accurately represents the assumptions of a given psychological theory.

In the domain of learning and retention, mathematical models are developed in order to represent how an individual retains knowledge based on the temporal aspects of a training schedule. Models often achieve this goal by representing three regularities of human memory: power law of learning, power law of decay, and the spacing effect. These three psychological phenomena have been represented in various mathematical models (Pavlik & Anderson, 2005; Raaijmakers, 2003; Walsh et al., 2018). The Predictive Performance Equation (PPE) is one particular mathematical model that has been found to account for a range of learning phenomena compared to other spacing effect models (Walsh, et al., 2018) and has been used to inform training applications . Each of these accomplishments was part of the explicit purpose of PPE's development, being used as a prescriptive educational tool.

However, despite the PPE's successful applications, its current formulation limits the estimation of psychological meaningful parameter estimates due to correlation across parameters. These limitations arise not because of the underlying psychological theory PPE represents or doubt of the empirical validity of the spacing phenomena, but because of PPE's chosen mathematical representation. In this paper, I review the current formulation of the PPE, address its limitations, and offer an alternative formulation of how they might be overcome.

Predictive Performance Equation

The PPE is composed of six individual equations, containing 4 free parameters. At the center of the PPE is the Activation term M_i (Eq. 1), which is a product of the learning term (N^c) and the forgetting term (T^{-d}). The learning term is on the unit of trial exposures (N) raised to a constant learning rate (c, usually .1). While the decay term is on the scale of model time (T), raised to a decay rate (d). From Eq. 1 it can be seen that M_i is on the scale of number of exposures and model time (T).

$$M_i = {N_i}^c * {T_i}^{-d} \quad \text{(Eq .1)}$$

A novel aspect of PPE is that model time (T) is modeled as a weighted average (Eq. 2) of time since all previous presentations of an item (Eq. 3). Thus, model time (T) is on the scale of the weighted average of wall clock time (often seconds).

$$T_i = \sum_{i=1}^{n-1} w_i * t_i \quad \text{(Eq .2)}$$

$$w_i = t_i^{-}x / \sum_{j=1}^{n-1} \frac{1}{t_j^{-}x} \quad \text{(Eq. 3)}$$

Additionally, PPE's decay rate (Eq. 5) dynamically changes over time, based on two free parameters, b and m, and the stability term. The stability term (Eq. 4) is a representation of the average natural logarithm of the lag_i of an item's history. Due to the fact that the natural log of the lag is taken, PPE's decay parameter is an unitless metric.

$$St_i = \left(\frac{1}{n-1} * \sum_{j-1}^{n-1} \frac{1}{ln\left(lag_j + e\right)}\right) \quad \text{(Eq. 4)}$$

$$d_i = b + m * St_i (Eq. 5)$$

Finally, to generate a prediction of performance, PPE's activation term (M_i) is nested within a logistic function (Eq. 6), which is controlled by two additional free parameters, τ and s, controlling the slope and intercept of the performance value. This formulation of activation value has been used in other learning contexts (Anderson, 2007)

$$P_i = \frac{1}{1+exp(\frac{\tau - M_i}{s})} \quad \text{(Eq. 6)}$$

Sources of correlation

As discussed in the previous section, PPE is composed of 6 equations with 4 free parameters. Equations 1 through 5 make up the PPE terms and Eq. 6 maps an unbounded activation term (M_i)onto a performance value. The free parameters are split such that they affect the PPE's decay (Eq. 5) term and the properties of the logistic function (Eq. 6). An unintended effect of this mathematical formulation is a high correlation between PPE's free parameters and an inability to compare best fitting parameters across individuals for psychometric evaluation (e.g., high versus low decay rates). The inability to compare parameters across participants is due to the fact that, since parameters correlate with each other these correlations must be taken into account before any parameter comparisons can be made across participants. Specifically, within the PPE this issue arises from two sources, (1) the PPE contains unbalanced units (i.e., M_i) and (2) the M_i term is nested within a logistic function. Each of these features have been shown in other psychological models to produce parameter correlation and issues with identifiability (Krefeld-Schwalb, Pachur, & Scheibeheen, *in press*). Here, we address the origin of both these sources in the PPE and propose an alternative formulation to remedy these correlation issues.

Unbalanced Units Unbalanced units refers to instances when particular terms within an equation are combined together without the units of those terms canceling out. For example, in PPE this occurs when computing the activation term M_i (Eq. 1) when the learning term is multiplied by the decay term. PPE's learning term is on the scale of instances of exposure (N_i), while PPE's forgetting term is on the scale of model time (T_i). Combining these two terms together, leads to an activation term M_i that is on the scale of number of events and model time, which results in highly correlated parameters, due to the fact that the free parameter (d_i) within each term are dependent on that term's scale. It is this correlation of parameters that hinders PPE's parameters being able to meaningfully represent individual differences within a sample due to the fact that any parameter estimate is dependent on that term's scale. The limitations of this formulation is not unique to PPE but has been found in other psychological models. Readers interested in a more thorough explanation should see Vincent and Steward (2020) and Stewart, Scheibehenne, and Pachur (2018).

Nested Equations A second source of intercorrelation within the PPE is the activation value (M_i) nested within a logistic function (Eq. 6), which is manipulated by its own free parameters (i.e., τ and s). The nesting structure creates three difficulties with model interpretation. First, nesting the activation term within the logistic equation allows for different M_i values to have equivalent performance values. Consequently, two people with identical learning and decay terms could be predicted to exhibit different performance in the future, which suggests that PPE's parameters (especially b and m) are difficult to interpret at face value (i.e., without also knowing the values of τ and s).. Second, within the logistic function, τ and s do not have independent effects on the activation term, allowing for multiple combinations of τ and s to create equivalent performance values. Third, having free parameters outside of the learning and forgetting terms obscures the interpretation of the PPE. Again, this issue is not unique to the PPE but has been noted as an issue with other psychological models Krefeld-Schwalb, Pachur, and Scheibeheen (in press).

Modified Predictive Performance Equation

As discussed above, PPE consists of unbalanced units and a nested equation, each of which lead to correlation across parameters. Both issues can be fixed by making relatively minor modifications to PPE's structure keeping the remaining assumptions of human performance intact. In the rest of the document We will refer to this new equation as the Modified PPE. In this section a proposed set of modifications to the PPE to reduce the correlation of parameters and improve the parsimony of the PPE are evaluated.

One cause of PPE's intercorrelations across the model's parameters is due to the M_i term being nested within a logistic equation (Eq. 5). This nesting step is required due to the fact that M_i is not bound between 0-1. However, this formulation is not required if both the learning and decay term are bound between 0-1, which would allow the learning and decay term to be combined together to estimate performance on a 0 - 1 scale (Eq. 7). To achieve this formulation, slight modifications are made to both the learning and decay term which are outlined here.

$$Performance_i = LearningTerm_i * DecayTerm_i \text{ (Eq. 7)}$$

Learning term In the standard PPE the learning term is produced as a power law. However, the reformulated learning term is exponential. Though there is a debate over the form of learning or forgetting term, the exponential formulation has been shown to better account for learning at an individual level performance over a power law (Heathcote, Brown, & Mewhort 2000). In this formulation the Modified PPE learning term (Eq. 8) has a learning rate m, which controls the rate at which material is acquired. When m is low information is acquired quickly, while when m increases the rate that individuals acquire information decreases. The benefit of this modified learning term is twofold. First, compared to the previous learning term (Eq. 1), it is now on a scale of 0-1. Second, due to the fact that N_i and m are used as exponents, which are unitless, the learning term can now be combined with the forgetting term .

$Learning\ Term_i = 1 - (e^{-N*(1-m)})$ (Eq. 8)

Decay Parameter One novel component of the PPE is that its decay term dynamically changes over time based on the temporal spacing of practice. This is a result of the assumption that spacing is a result of attention moderating the spacing effect (Walsh et al. 2018). The Modified PPE retains the same assumption of the use of the stability term (Eq.4),which accounts for an item's previous temporal history of presentations. The addition of the lag term $\frac{1}{log\ (lag_i)}$, which represents the most recent lag between exposures was added to PPE's decay parameter due to the fact that M_i is no longer nested in the logistics term and augmented by the τ and s parameter. The lag term is modified by the same learning rate (m) parameter as used in the learning term (Eq.8). When learning rate is low, the effect of the most recent lag is minimized, while when it is high the effect of the most recent lag is maximized. The stability term is manipulated according to the b parameter, which controls the effects that the previous temporal schedule (i.e, spaced practice) has on performance. This subtraction of the maximum value is used to format the decay term serving as a decay intercept based on the largest decay within a set of practice items (Eq. 9). The benefit of this decay term is that it is now composed of two terms (lag_i and St_i) that each represent separate aspects of the performance and each manipulated by their own free parameter.

$d = (\frac{1}{log(lag_i)} * m + St_i\ * b) - max(\frac{1}{log(lag_i)} * m + St_i\ * b))$ (Eq. 9)

Forgetting Term For the decay term, the standard power law formulation was retained from the PPE. However, the power law is expressed as a ratio which allows the forgetting term to be expressed as a unitless metric between 0-1 (Eq. 10).

$$Forgetting\ Term = \frac{T_i^{-d_i}}{1+T_i^{-d_i}} \quad \text{(Eq.10)}$$

Summary of Changes to the PPE

Here we reviewed the PPE, a model of learning and retention which has shown great promise in accounting for both laboratory and real world findings. However, features within the PPE lead to correlation across parameters and hinder it from being used to estimate psychological constructs (i.e., learning and decay rates). To correct these limitations, we have proposed a new formulation of the PPE, decreasing the number of free parameters while retaining PPE's unique features: multiplicative performance, model time, and variable decay term.

Method

To highlight a comparison between the standard and modified PPE, a model comparison was conducted, highlighting the correlation across free parameters and the benefits of the PPE' formulation.

Participants Sixty-one participants were recruited from a midwestern university in this paired-associate learning study. All participants completed a total of three experimental sessions spanning a three-week period.

Task Stimuli Over the course of the experiment participants memorized a set of 30 Japanese-English words. All of the words used in this study were taken from the Medical Research Council (MRC) Psycholinguistic Database manual and have been used in other previous memory studies (e.g., Pavlik & Anderson, 2005).

Experimental Design and Procedure During the experiment, an item's training schedule was manipulated according to inter-session interval (ISI) and inter-trial interval (ITI) over the course of experimental sessions. The ISI controlled the amount of time between the 1st and 2nd experimental session, with a fixed 7 day ISI between the 2nd and 3rd session across all conditions.. The ISIs in this study were fixed at short (5 min), medium (7 days), and long (14 days) delay. The ITI manipulated the number of trials between presentations of the same item within a session. Two ITIs consisting of a short (every 2 trials) and long (every 11 trials) delay were embedded in each experimental session.
During the study, participants, with no knowledge of the Japanese language, were given instructions for the paired associate learning task and had an opportunity to ask any questions. Once participants began the experiment, they were shown a Japanese word (e.g., "kanboku") on the screen and asked to type the English translation (e.g., "bush") to the Japanese word. Upon first presentation of a word, participants were shown the English translation and asked to type the correct answer to ensure the item was studied. During all subsequent presentations, participants were asked to recall and type the English translation from memory. Participants were given a maximum of 7 seconds to type their answer during each trial. If a participant could not generate a response within 7 seconds, then their answer was considered incorrect. At the end of each trial participants were given feedback (correct or incorrect) and given 2 seconds to study the correct answer.

Bayesian Models

To examine the two implementations of the PPE, Bayesian hierarchical models of both the Standard and Modified PPE were implemented in JAGS (Plummer, 2012). Each model was run with 3 MCMC chains, run for 9000 iterations, with a fixed burn in period of 1000 iterations. Each models priors were chosen so that the prior predictions from each model expressed the standard learning phenomena expected from the learning schedule (i.e., slower learning in the long vs short ITI condition with more decay between the sessions in the short compared to the long ITI). Each model was fit to each of the Japanese-English word pairs across the three experimental sessions.

Results

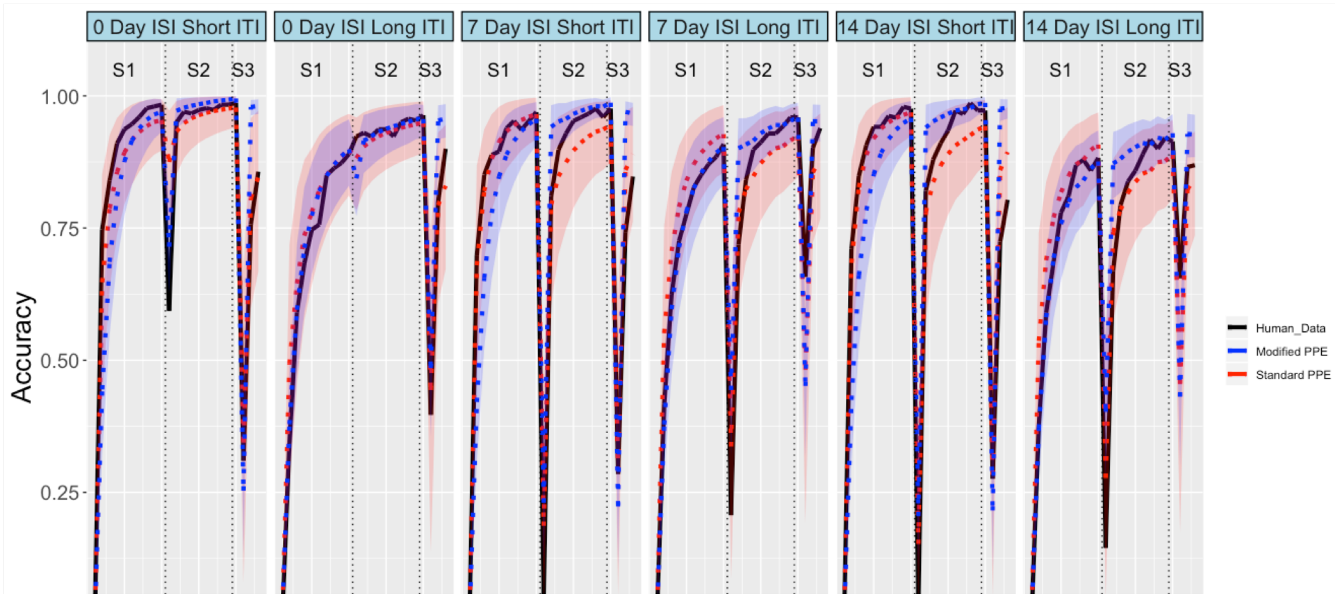

Figure 1. The average +/- 95% HDI posterior model fit of the Standard (dashed red line and ribbon) and Modified (dashed blue line and ribbon) PPE to the participants' performance (sold black line) six experimental conditions varying the three inter session interval (0, 7, and 14 days) and the two inter trial intervals (short and long).

To compare the Standard and Modified PPE equation, a comparison across three different metrics was performed. First, we examined how well each model fit to the performance of subjects across each of the learning schedules. Second, we compared the correlation between each model's free parameters. Third, the relationship between the participants' parameter estimates and learning retention were examined.

Model Fit

We first examined the average performance of participants and each model's posterior performance estimates across the six different learning schedules. An examination of the models' average fit to the participants performance reveals several interesting qualitative findings (Figure 1). First, the Modified PPE for the most part has much narrower 95% HDI compared to the Standard PPE. This difference in precision between the two models is the result of the difference in complexity. The Standard PPE has 4 free parameters, with the activation term (M_i) being nested within a logistic equation, which gives the model additional flexibility. An example of this additional flexibility can be seen in the relearning between in 2nd and 3rd experimental sessions: the Standard PPE shows quick but attenuated relearning across sessions, while the Modified PPE shows quick relearning between sessions.

Table 1. The correlation (r) and root mean squared deviation (RMSD) between the standard and modified PPE across each of the six learning schedules.

		Standard PPE		Modified PPE	
ISI	ITI	*r*	*RMSD*	*r*	*RMSD*
0	Short	0.96	0.07	0.93	0.10
0	Long	0.99	0.04	0.98	0.05
7	Short	0.99	0.05	0.93	0.11
7	Long	0.96	0.07	0.95	0.08
14	Short	0.99	0.05	0.92	0.11
14	Long	0.96	0.07	0.93	0.09

To evaluate the fit of both models to the participants' performance across the three experimental sessions, the correlation (r) and root mean squared deviation ($RMSD$) between the average accuracy and each model's posterior performance were calculated (Table 1). Both models fit the average performance of participants across all of the experimental conditions quite well, with the Standard PPE having a slightly higher correlation and lower RMSD compared to the Modified PPE. However, a Bayes factor found the the Modified PPE to be strongly preferred to the Standard PPE (BF > 30)These results suggest strong evidence in favor of the Modified over the Standard PPE, suggesting that the Modified PPE is a more parsimonious model compared to the Standard PPE in this context.

Parameter Intercorrelation

Next, we evaluated the intercorrelations between each of the models' free parameters (Figure 2). A correlation between two parameters reveals a functional interdependence, which hinders theoretical interpretations of the parameters.

Standard PPE To evaluate the comparison between the Standard PPE's free parameters, the correlation between all free parameters (b, m, s, τ) were calculated from the models

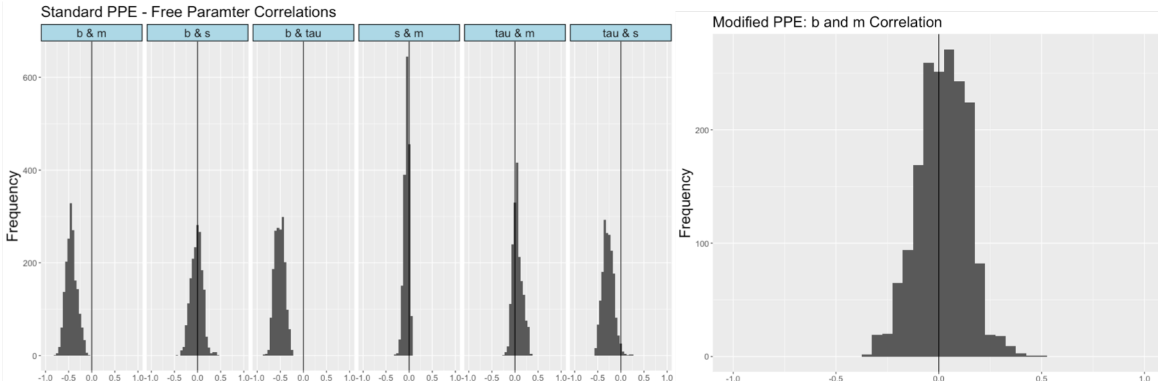

Figure 2. A histogram of each the free parameters of the Standard (left panel) and the Modified PPE (right panel) correlation with each other for each item presented over the course of the experiment.

fit for each of the Japanese – English word pairs studied by participants (Figure 2 – left panel). The results of the correlation between the Standard PPE's free parameters are apparent and as expected from the analysis of the Standard PPE formulation. First, there was a moderate negative correlation between the *b* and *m* parameter in the standard PPE decay equation (Eq. 5). This correlation between the *b* and *m* parameters occurs due to the fact the decay parameter is structured as a linear regression with the product of the stability term and *m* being added to the *b* parameter. From this construction the same decay value can be achieved under a variety of *b* and *m* combinations. Second, the *b* parameter is seen to negatively correlate with τ parameter (Figure 2 – left panel). This correlation is caused by nesting the Standard PPE's activation term (M_i) within the logistic term (Eq. 6). Due to the fact that τ affects the Standard PPE's performance estimation outside of the activation term additional variance in the participants performance can be explained by manipulating either the τ or *b* parameter. Finally, a smaller negative correlation between the τ and *s* parameter was observed. This correlation is the result of the structure of the logistic term and unbalanced units of the Standard PPE's activation term subtracted by the τ.

Modified PPE Compared to the standard PPE, the Modified PPE has only two parameters: *b* and *m.* Overall, the correlation between the *b* and *m* parameter is minimal compared to some of the correlations across parameters that were observed in the Standard PPE. This reduced correlation is the result of removing the logistic term from the equation, thus not having any nested terms within the equation and making sure the *b* and *m* parameters each affect only one term in the decay parameter, the stability term and the lag. The effect of each of these manipulations is that the Modified PPE is greatly simplified.

Measuring Aspects of Performance

In our final comparison between the Standard and Modified PPE, each model's subject-level free parameters and specific aspects of the participants' performance were examined. The two relevant aspects of the participants' performance chosen for this paper were the participants' overall accuracy and retention between sessions (i.e. accuracy on the 1st trial during the 2nd and 3rd session trials 11 & 21). Ideally, a model's free parameters should represent a latent theoretical construct, such as learning and decay, which then map on to particular measures of behavior.

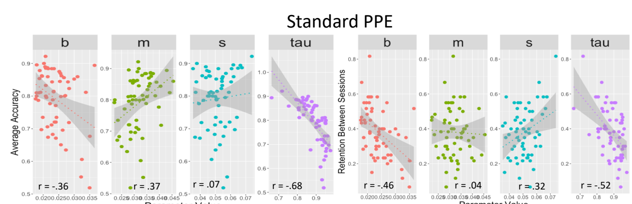

Figure 3. Scatter plot between participants' overall average performance (left plot) and the average performance during the 1st trial during the 2nd and 3rd session and the free parameters (right plot) from the Standard PPE (left four columns).

Standard PPE As seen in Figure 3, the Standard PPE's free subject-level parameters were seen to have a moderate correlation with both the participant's overall average performance and their retention between sessions. However, with both measures τ was found to have the highest correlation with both overall accuracy and the retention between sessions, compared to both the *b, m,* and *s* parameter. This result highlights that the τ parameter has a disproportionate influence on the Standard PPE's performance estimates. The predominant influence of τ can be seen as problematic due to the fact that τ modifies the activation term (M_i) and does not have any direct influence on either model time (T_i) or the stability term (St_i). One potential cause for the limited influence of the *b* and *m* parameter on accounting for performance, is the correlation between each other (i.e., *b* & *m*) and the τ parameter.

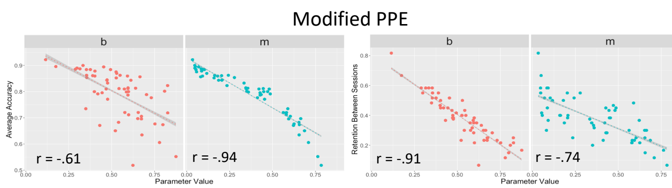

Figure 4. Scatter plot between participants' overall average performance (left plot) and the average performance during the 1st trial during the 2nd and 3rd session and the free parameters (right plot) Modified PPE.

Modified PPE: In contrast to the Standard PPE, the free parameters in the Modified PPE are both seen to have a strong relationship with both the participants' overall accuracy and retention between sessions (Figure 4 - Panel B). Further, the degree of each parameter's relationship can be understood from a theoretical perspective. The *m* parameter is accounting for the participants' ability to learn the Japanese – English word pairs during the experiment. This relationship is in line with the *m* parameter's function within the Modified PPE, affecting the rate at which information is learned. The *b* parameter, on the other hand, which affects the model's stability term, is seen to predominantly account for the participants' retention between sessions. These results highlight another benefit of the Modified over the Standard PPE. The simplified structure of the Modified PPE allows for the parameters to better summarize and map on to particular aspects of the participants' performance.

Discussion

In this paper, the formulation of the Standard PPE was examined. Two aspects of the PPE's structure were identified as contributing to the correlation between PPE's free parameters. Correlation between parameters increases a model's complexity and obscures the meaning that can be attributed to particular parameter estimates.

To reduce this intercorrelation across parameters, several modifications were made to the Standard PPE's structure, removing the activation term (M_i) from the logistic equation and modifying the forgetting and learning terms. Although these modifications to the Standard PPE changed the structure of the equations, the components unique to the PPE relative to other spacing models (i.e., variable decay rate, stability term, and model time) remained intact. A formal model comparison between the Standard and Modified PPE revealed that the Modified PPE was able to (1) account for the participants' performance across the three experimental sessions, (2) greatly reduced the correlation across its two free parameters, and (3) parameter estimates mapped on to specific aspects of the participants' performance.

It is important to note that the results reported in this paper do not invalidate any previous findings of the PPE, but simply address the meaning that can be attributed to its parameter estimates. PPE was initially developed as a predictive tool and to meet a set of applied criteria (i.e., assign prescriptive scheduling, calibrate quickly to prior performance, account for relearning of spaced items after a delay; see Walsh et al., 2018 for full list). Along these criteria the Standard PPE has succeeded and has been used successfully as a predictive tool across different applied domains.

Attempting to explain data from a theoretical point of view and predicting new observations are opposing goals for scientific models (Shmueli, 2010) and neither one should be considered superior to the other. Instead, a balance between these two extremes should be found based on the pragmatic goals of the research question. If the goal is to use the PPE as a method to predict future learning and retention behavior of an individual, then the Standard PPE's formulation is acceptable. In contrast, if the goal is to summarize an individual's performance in terms of psychologically latent values (i.e. decay, learning) or to compare the best-fitting parameters across individuals to evaluate individual differences, then the Modified PPE proposed here is a more appropriate tool.

Several limitations need to be addressed within this paper. First, additional research needs to be conducted to further explore how well the Modified PPE can account for performance across longer and more variable learning schedules relative to the Standard PPE. Here, the Standard and Modified PPE were compared across only six unique learning conditions. Future research should compare the two models along a variety of both spaced and non-spaced learning schedules to better find where these two models differ. Second, this paper focused on reducing the parameter correlation across the Modified PPE parameters, to simplify the model and reduce its dimensionality. Future research should explore using Modified PPE for psychometric purposes, evaluating if either of its parameters correlate with particular psychological constructs such as working memory or attention.

Conclusion Mathematical models of psychological theories are useful tools for theory evaluation, development and applied technologies. For these goals to be met, care should be taken to ensure that a model's formulation and representation are adequate and are in line with their verbal descriptions. By attending to how particular implementations of theories are represented, a balance between mathematical, statistical, and scientific validity can be found.

Acknowledgments

The authors would like to thank Joe Houpt, Christopher Fisher, and Noah Thomas for helpful discussions and comments on this work. This research was supported by MGC's participation through an appointment to the Oak Ridge Institute for Science and Education (ORISE) Student Research Participation Program. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government

References

Krefeld-Schwalb, A., Scheibehenne, B., & Pachur, T. (2019, February 1). Structural Parameter Interdependencies in Cognitive Models. https://doi.org/10.31234/osf.io/pxmnw

Heathcote, A., Brown, S., & Mewhort, D. J. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic bulletin & review*, *7*(2), 185-207.

Pavlik Jr, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559-586.

Plummer, M. (2012). JAGS Version 3.3. 0 user manual. *International Agency for Research on Cancer, Lyon, France*.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431-452.

Shmueli, G. (2010). To explain or to predict?. *Statistical science*, *25*(3), 289-310.

Vincent, B. T., & Stewart, N. (2020). The case of muddled units in temporal discounting. *Cognition*, *198*, 104203.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., Krusmark, M., Myung, J. I., ... & Zhou, R. (2018). Mechanisms underlying the spacing effect in learning: A comparison of three computational models. *Journal of Experimental Psychology: General*, *147*(9), 1325.