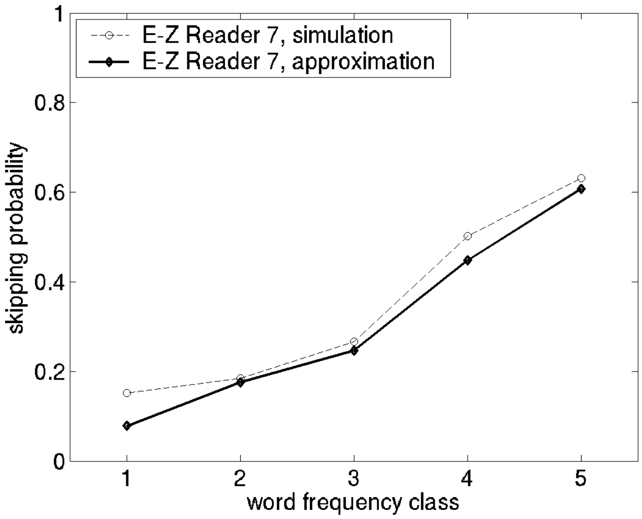

Figure 1 (Engbert & Kliegl). Skipping probability as a function of word frequency class.

msec, 90.8 msec, 60.7 msec, and 44.4 msec for L_1, using Equation 3 and corresponding parameter values given in Reichle et al. The average value of L_2 corresponds to an arbitrary word (the word left of the skipped word). Therefore, we used the ensemble average of L_2 over all words the corpus of sentences,[2] denoted by $\langle L_2 \rangle$ = 82.3 msec. For a gamma distribution of order n=8 and a mean labile saccade duration M_1=187 msec, we obtained τ =20.8 msec. The resulting estimates for the skipping probability $p_{\text{E-Z Reader}}$ are in good agreement with simulated data from the target article (see Fig. 1).

In SWIFT, a field of lexical activities $a_n(t)$ evolves over time. The probability of target selection is given by the relative lexical activity. As a consequence, no additional assumptions must be made to produce forward saccades, refixations, and regressions. The probability of skipping word$_{n+1}$ is given by the probability to select word$_{n+2}$ as the next saccade target, which is computed by the fraction

$$p_{SWIFT} = \left. \frac{a_{n+2}(t)}{\sum_{k=1}^{n+2} a_k(t)} \right|_{t = \text{target selection}} \tag{3}$$

There is no oculomotor contribution to the skipping probability in Eq. (3) – an important difference to Equation (2) for E-Z Reader. Numerical estimates for p_{SWIFT} can be obtained by evaluating the set of lexical activities at the point in time where target selection occurs in SWIFT (for details see Engbert et al. 2002).

Diverging predictions can be derived from SAS and SWIFT models. In E-Z Reader, the probability of word skipping will depend on oculomotor parameters, because of the competition between saccade programming and word identification. In SWIFT, the competition between words for becoming selected as the next saccade target implies a structural stability of word skipping against oculomotor parameters. Therefore, dynamic models generate highly specific predictions, which might be most stimulating for future research: The current controversy on mechanisms of eye-movement control will still be resolvable by experimental results.

ACKNOWLEDGMENTS

We thank Erik D. Reichle for providing us with the simulated data for E-Z Reader 7. This work was supported by Deutsche Forschungsgemeinschaft (grants KL-955/3).

NOTES

1. The gamma distribution for saccade latencies can be written as $q_\tau^n = \frac{1}{\tau - n!}\left(\frac{t}{\tau}\right)^n \exp\left(-\frac{t}{\tau}\right)$, where τ is a time constant and n is the order of the distribution. Mean value and standard deviation are given by $\mu = (n+1)\tau$ and $\sigma = \sqrt{n+1}\tau$. For a relation of standard deviation to mean of one third (Reichle et al. 1998), we have to choose a gamma distribution of order n = 8.

2. This procedure may be interpreted as a *mean field approximation*, that is, using the average processing difficulty of the word left to the skipped word. To compute L_1 and $\langle L_2 \rangle$ according to Equation 3 in the target article, we used word frequencies, predictabilities, and the parameters β_1, β_2, and Δ.

Throwing the baby out with the bathwater: Problems in modeling aggregated eye-movement data

Gary Feng

Department of Psychology: Social and Health Sciences, Duke University, Durham, NC 27708. **garyfeng@duke.edu**
http://www.psych-shs.duke.edu/faculty/facultyfeng.html

Abstract: Parameters in E-Z Reader models are estimated on the basis of a simple data set consisting of 30 means. Because of heavy aggregation, the data have a severe problem of multicolinearity and are unable to adequately constrain parameter values. This could give the model more power than the empirical data warrant. Future models should exploit the richness of eye movement data and avoid excessive aggregation.

Because eye movement techniques produce an enormous amount of raw data, it is common practice to perform some sort of data reduction prior to modeling. However, there is a delicate balance between reducing computational complexity and preserving meaningful variance in the data. The E-Z Reader 7 model and its predecessors (Reichle et al. 1998; 1999) posit a comprehensive and elegant set of eye-movement control mechanisms, but the data set used to fit the models is too impoverished to adequately test these models.

The empirical data for E-Z Reader models (from Schilling et al. 1998) are averages of six eye movement variables (single fixation duration, first fixation duration, gaze duration, and the probability of skipping, making single fixations, and making two fixations) over five word frequency levels (Reichle et al. 1998, Table 1; see also Note 6 of the target article). Unfortunately, the structure of the empirical data is ill formed. The six variables are so highly correlated that the data space has far fewer than six independent dimensions, a problem known as *multicolinearity* in linear regression analysis.

Pair-wise correlation coefficients among the six variables range from 0.85 (between skipping rate and first fixation duration) to 0.998 (between first fixation duration and single fixation duration). Furthermore, all eye-movement measures are highly correlated with the logarithm of word frequency. A principal component analysis showed that the first component accounts for 94.6% of the total variance, the first two components account for 98.6%, and the first three components account for 99.999% of total variance. In short, with only 5% loss of information, the six eye-movement variables can be effectively reduced to a single variable, which in turn has an almost perfect linear relationship with log-transformed word frequency.

The consequences of multicolinearity in the dataset are profound. Free parameters (ranging from five in E-Z Reader 1 to at least seven in E-Z Reader 7 models) were effectively estimated on the basis of only five data points, creating a classic identification problem in parameter estimation, where some parameter values may be varied freely without affecting model fit. Moreover, flaws in the data threaten the internal validity of E-Z Reader as an empirical model. If we believe in the principle of parsimony, then the only model that will survive Occam's Razor would be something like "any eye-movement measure is a linear function of the log-transformed word frequency," which is both uninformative and wrong (see Kliegl et al. 1982).

Is the problem of multicolinearity confined to the Schilling et al. (1998) data set? The answer is no. A similar analysis based on our own data (Feng et al. 2003) shows the same pattern of multicolinearity. The problem stems from two sources.

1. Composite eye-movement measures (see Inhoff & Radach 1998), such as gaze duration and probability of skipping, are statistics computed from individual fixations. Because these statistics are calculated on the same sample of fixations, moderate to strong correlations are expected among them. For example, the fixations counted toward single-fixation duration are a subset of first-fixation duration, which is in turn a subset of gaze duration.

2. These correlations are further concentrated as raw data are aggregated to get a "clean" picture suitable for modeling. For example, in our adult reading data, the correlation between first fixation duration and gaze duration is 0.71 when the unit of analysis is per subject per word (N=24,089). It becomes 0.80 when we average across subjects (N=3,599 words), and 0.95 if we only consider five word frequency levels (N=5) and average across both subjects and words.

As long as only a few means of composite eye-movement variables are used for modeling, the problem of multicolinearity will be unavoidable. Therefore, ingenious and intricate theories such as E-Z Reader will remain untestable. The only solution to this problem is to reinstate the richness of the eye-movement data for modeling. There are at least three approaches to this end:

1. Use less aggregated data.
2. Model distribution functions of eye movement variables (e.g., Feng et al. 2001; 2003; McConkie & Dyre 2000).
3. Use raw data instead of composite eye-movement measures (Feng 2001).

In addition to the multicolinearity problem, there are several important flaws in the parameter estimation procedure shared by all E-Z Reader models (see Reichle et al. 1998, p. 157). Instead of normalizing the difference between model predictions and observed values, the authors erroneously squared the difference. Consequently, fixation duration variables, which have a much larger scale than do probability variables, contributed approximately 100 times as much to the index of model fit as did the probability variables (estimation based on Reichle et al. 1998, Table 1). Another error is the use of the standard deviation in the normalization. Because the comparisons were between observed means and simulated means, the sample standard error should be used in the denominator (see Hayes 1988). As a result, the goodness-of-fit index was shrunk by a factor of the square root of *N*. Finally, it is disappointing that there was no attempt to test the fit of each model statistically, or statistically compare successive models. Further analyses on the impact of these factors can be found in Feng (2001).

It may seem paradoxical that even though it has serious problems in parameter estimation, E-Z Reader 7 is successful in simulating many well-known reading eye-movement phenomena. A possible explanation is that precisely because the impoverished empirical data could not provide adequate constraints over parameter values, the authors gained more freedom in assigning parameter values that maximize simulation performance. This would predict that the model's simulation performance would be hampered if the data contain more information, something that could be empirically tested.

In summary, the problems discussed here – multicolinearity in data and issues with parameter estimation and model testing – are fairly low-level. However, a model is ultimately only as good as the data and algorithms on which it is based. There is not enough evidence to conclude that E-Z Reader is empirically validated. Nonetheless, we should not throw out the model with the statistical problems. These issues are not difficult to fix. I look forward to seeing an E-Z Reader 8 that is on a solid statistical footing. Meanwhile, future modeling work should fully exploit the richness of reading eye movements and be wary of the limitations of aggregated data.

ACKNOWLEDGMENTS
I am grateful to Kevin Miller and Ruth Day for helpful comments on early drafts.

Serial programming for saccades: Does it all add up?

John M. Findlay and Sarah J. White

Centre for Vision and Visual Cognition, Department of Psychology, University of Durham, Durham, DH1 3LE, United Kingdom.
j.m.findlay@durham.ac.uk s.j.white@durham.ac.uk
http://psychology.dur.ac.uk/staff/details.asp?name=FINDLAY
http://www.dur.ac.uk/s.j.white/

Abstract: This commentary analyses the quantitative parameters of Reichle et al.'s model, using estimates when explicit information is not provided. The analysis highlights certain features that appear to be necessary to make the model work and ends by noting a possible problem concerning the variability associated with oculomotor programming.

Reichle et al.'s target article presents a model of eye control during reading that is impressive in a number of ways. It is fully explicit, quantitative, and economical, and it brings in known features of the visual system (differential magnification) and oculomotor system. It provides a good account of a number of observed phenomena and a quantitative fit to data. Its appearance in *BBS* is particularly welcome because if it proves robust against criticism, this must be regarded as a triumph not only for the model itself but also for the serial stage approach to modeling that underpins it.

The duration of fixations are modeled on the basis of a signal traveling through a number of stages that are strictly serial with the visual, lexical, and oculomotor processes taking place sequentially. These stages are shown in Figures 3 and 13 of the target article; and the latter figure in particular suggests that the time-consuming processes leading to saccades are conceived as the time for signals to traverse brain regions. This represents a different tradition and philosophy to the approach of Findlay and Walker (1999), where the emphasis was on specific time-consuming processes of competitive inhibition, particularly in the late oculomotor stages. Some common ground might be found in the separation of the programming of saccade amplitude from the remainder of the programming. This occurs through the direct (dashed line of Fig. 3) pathway from the early stage of visual processing bypassing the word identification system. Section 3.1.3 indicates that this pathway provides the low spatial frequency information needed to program a saccade. However, it would appear that there also needs to be a modulatory influence from the word segmentation process on this pathway, since the whole basis of the model is that saccades are programmed to words.

The remainder of this commentary works through the model in detail, following the commentators' understanding and looking particularly at the time course of events.

The seriality has the consequence that the duration of a fixation can be expressed as a sum of contributions from the component stages

$$\text{FXDUR} = t(V)+L_1+M_1+M_2\text{-}(OV_V+OV_L+OV_M) \qquad (1)$$

where $t(V)$, L_1, M_1, and M_2 are defined as in the model. OV_V, OV_L, and OV_M are introduced to denote the modifications that occur when the model is working dynamically, since overlap (OV) processes can occur. OV_V and OV_L are savings in the visual and lexical stages, respectively, that can come from peripheral preview. OV_M reflects changes in oculomotor preparation time when saccadic programming stages overlap. All the components are described clearly in the target article and, although the detailed magnitudes can be made only with precise knowledge of the text being read, it is possible to make estimations of the distributions. The