Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 1, 219–229

0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.1.219

BRIEF REPORTS

Modeling Distributions of Immediate Memory Effects: No Strategies Needed?

C. Philip Beaman
University of Reading

Ian Neath and Aimée M. Surprenant
Memorial University of Newfoundland

Many models of immediate memory predict the presence or absence of various effects, but none have been tested to see whether they predict an appropriate distribution of effect sizes. The authors show that the feature model (J. S. Nairne, 1990) produces appropriate distributions of effect sizes for both the phonological confusion effect and the word-length effect. The model produces the appropriate number of reversals, when participants are more accurate with similar items or long items, and also correctly predicts that participants performing less well overall demonstrate smaller and less reliable phonological similarity and word-length effects and are more likely to show reversals. These patterns appear within the model without the need to assume a change in encoding or rehearsal strategy or the deployment of a different storage buffer. The implications of these results and the wider applicability of the distribution-modeling approach are discussed.

Keywords: memory modeling, individual differences, immediate recall

A number of phenomena are assumed to typify the functioning of immediate memory, including the acoustic confusion or phonological similarity effect (PSE; reduced recall levels for similar sounding items; Conrad, 1964) and the word-length effect (WLE; reduced recall for words with more syllables; Baddeley, Thomson, & Buchanan, 1975). The PSE, for example, is frequently taken to be the "signature" that indicates the involvement of the phonological store component of the working memory model (although, see Jones, Hughes, & Macken, 2007). The WLE is often associated with the operation of an articulatory rehearsal loop (although, see Lovatt & Avons, 2001; Neath, Bireta, & Surprenant, 2003). Among patient populations, the absence of such effects at the individual level is used as evidence for impairment of particular short-term memory functions (Silveri, Di Betta, Filippini, Leggio, & Molinari, 1998; Vallar & Shallice, 1990). Thus, a reduced PSE is often assumed to reflect the lack of a phonological storage strategy or (in neuropsychological patients) damage to the phonological store. Similarly, a reduced WLE is easily interpreted (e.g., among young children and older adults) as reflecting poor articulatory rehearsal or, again among neuropsychological patients, damage to the articulatory loop.

C. Philip Beaman, Department of Psychology, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom; Ian Neath and Aimée M. Surprenant, Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

Portions of this research were carried out while Ian Neath and Aimée M. Surprenant were at Purdue University and also while they were visiting professors at City University, London.

Correspondence concerning this article should be addressed to C. Philip Beaman, Department of Psychology, School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Whiteknights, Reading RG6 6AL, United Kingdom. E-mail: c.p.beaman@reading.ac.uk

As noted by Logie and colleagues (Della Sala & Logie, 1997; Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996), both PSEs and WLEs are absent—or even reversed—in the recall protocols of some normally functioning participants, suggesting that although these effects occur in the aggregate or group case traditionally studied within cognitive psychology, they are not necessarily present at the individual level. One purpose of the current work is to determine the distribution of effect sizes observed when standard PSE and WLE experiment methodologies are used. A second purpose is to assess how well the feature model (Nairne, 1990) accounts for the observed interindividual variation. Although Melton (1967) noted that a theory of memory must, perforce, include a theory of individual differences in the processes specified by the theory, few models of immediate memory actually explore these processes in any detail (Lewandowsky & Heit, 2006).

The magnitude of some effects could vary among the population for theoretically interesting reasons, and any candidate model of the effects in question should track this variation. Both effects have been linked to particular strategies. For example, it has been suggested that the perceived difficulty of an immediate verbal memory task might result in the abandonment of a phonological code in favor of other, nonphonological, coding mechanisms (see the debate between Baddeley & Larsen, 2007, and Jones et al., 2007). Similarly, participants who more effectively use rehearsal might show both superior overall performance and a larger effect of word length than participants who do not use rehearsal, a line of argument used in the developmental literature to differentiate the performance of younger children from that of older children and adults (Hitch & Halliday, 1983).

A previous study (Logie et al., 1996) reported reliable interindividual differences in PSEs and WLEs using a span technique to establish the incidence of variation within a normal population and hence inform interpretation of neuropsychological patient data. However, span scoring is less sensitive than serial recall or reconstruction, which record multiple errors per trial compared with span scoring's all-or-none system. Additionally, Logie et al. (1996) interpreted their data in terms of a variety of strategy choices among their participants, on the basis of postexperimental questioning. There are a number of reasons to question this interpretation. First, Logie et al. provided no indication of how strategies were identified and classified, nor did they provide any indication of interrater reliability in classifying strategies. Second, post hoc introspective accounts are subject to a number of biases and must be interpreted cautiously. Third, even if strategies are both reported and classified correctly, covariation of strategy and effect size constitutes a correlational, not necessarily a causal, relationship. Fourth, when span was included as a covariate, the "effect" of strategies reported by Logie et al. drops considerably. A factor other than strategy choice must therefore be responsible for the majority of the variation reported. Finally, although not formally analyzed, Logie et al.'s results show a distribution of effect sizes that appears normal, or close to normal, not the bimodal distribution one might expect from a mix of phonological and nonphonological encoding strategies. Given these issues, as well as more recent theoretical work questioning the role of these strategies (Hulme et al., 2006; Hulme, Surprenant, Bireta, Stuart, & Neath, 2004; Nairne, 2002), we examine whether a nonstrategy account of interindividual variation provides a viable model of the data. This approach parallels other modeling studies showing that strategy as an explanatory construct may not be required as frequently as often assumed (Brown & Hulme, 1995; Neath & Nairne, 1995).

The Feature Model

Space precludes a detailed account of the feature model; the implementation was identical to that used in other simulations (Beaman, 2006; Neath, 1999, 2000; Neath & Nairne, 1995; Neath & Surprenant, 2007). Here, we provide only an overview. The basic idea is that recall is guided by a set of cues in primary memory (PM), which are either more or less effective in identifying the target item or event from a search set defined within secondary memory (SM). Cues do not decay but are subject to interference. Anything that makes the cues less effective (e.g., similarity between the representations of to-be-recalled targets) or degrades the cues such that two or more targets could fit the profile specified by the cues renders recall less accurate.

Formally, items in SM and cues in PM are presumed to comprise vectors of modality-independent and modality-dependent features. For simulation purposes, feature values are randomly generated. The absolute number of features is largely irrelevant in determining the major properties of the model (see, e.g., Figure 4.8 of Neath & Surprenant, 2003). The main source of interference in the model is feature overwriting: If feature x of item $n + 1$ is the same as feature x of item n, then feature x of item n is lost, or overwritten, and returns a value of 0.

The relative number of accurate features available to cue the item in SM dictates recall performance. The difference between the target SM item and its PM cue is calculated according to Equation 1. The value M_k is equal to 1 if the feature at position k of PM cue i does not match the feature at the corresponding position of SM representation j and 0 otherwise. The number of mismatches across the features is summed in the numerator of Equation 1. The value N is the number of features in each of the vectors, and a is a scaling parameter representing overall level of attention.

$$d_{ij} = a\Sigma \frac{M_k}{N}. \tag{1}$$

The difference between the PM and SM items is transformed to provide a similarity metric (Equation 2).

$$s(i,j) = e^{-d_{ij}}. \tag{2}$$

The probability that a particular secondary memory trace SM_j will be "sampled" given a particular primary memory cue PM_i is given by a similarity-based choice rule (Equation 3).

$$P(SM_j|PM_i) = \frac{s(i,j)}{\sum_{l=1}^{n} s(i,l)}. \tag{3}$$

The probability of recovering a sampled item is then given by Equation 4, where c is a constant and r is the number of times the sampled item has already been recalled on this trial. This equation and the r parameter are used to reduce the likelihood of recalling the same item on multiple occasions, which participants avoid doing even when the same item is repeated within the to-be-recalled list.

$$P_r = e^{-cr}. \tag{4}$$

In the feature model, order information is represented as a point in multidimensional space and this point can perturb (or drift) along the relevant dimension as described by Estes's (1972) equations. The probability that a cue's encoded representation will perturb along the position dimension during a particular time interval is given by the parameter θ (which is held constant at .05); perturbations are equally likely in either direction. The probability that an item, I, will occupy a particular position, p, during the next time interval, $t + 1$, is given by Equation 5:

$$I_{p,t+1} = (1 - \theta)I_{p,t} + \left(\frac{\theta}{2}\right)I_{p-1,t} + \left(\frac{\theta}{2}\right)\mathrm{I}_{\mathrm{p}+1,t}. \tag{5}$$

For the first and last positions, a slightly different equation is used to reflect the assumption that an item cannot perturb to a position outside the list. For the first position,

$$I_{1,t+1} = \left(1 - \frac{\theta}{2}\right)\mathrm{I}_{1,t} + \left(\frac{\theta}{2}\right)I_{2,t}. \tag{6}$$

For the last position, n,

$$I_{n,t+1} = \left(1 - \frac{\theta}{2}\right)I_{n,t} + \left(\frac{\theta}{2}\right)I_{n-1,t}. \tag{7}$$

Recall begins by determining, for each cue in PM, which was most likely in Position 1 originally. To recall the second item, the model uses the cue that was most likely in Position 2, and so on (see Neath, 1999, for more details).

The feature model produces PSEs primarily through Equation 3. Phonologically similar items are represented by vectors that have a higher number of features (both modality independent and modality dependent) with the same value. This has the net effect of increasing the value of the denominator and thus reducing the probability of retrieving the appropriate item (see Nairne, 1990).

The WLE is explained by assuming that the more complex an item, the more segments it contains, and therefore the more chance there is of an error (see Neath & Nairne, 1995). Long words are typically more complex than short words and therefore have more segments, and so there is more chance of making an error when assembling the segments for output (Caplan, Rochon, & Waters, 1992). If a segment error occurs, there is a loss of feature information, implemented by setting half of the modality-independent features to 0 values (see Neath & Nairne, 1995, for full details).

The feature model was chosen to model distributions of PSEs and WLEs for three reasons. First, it has been shown to account well for both the main PSE and WLE data. Second, the model as currently implemented does not include rehearsal or any other more complex strategy. Any predictions of the model must therefore be strategy independent and not reliant on the use of a particular maintenance or rehearsal strategy. Third, the model lends itself to a type of simulation previously used by Hintzman (1986, 1988) and more recently used by Cooper, Schwartz, Yule, and Shallice (2005) in which one run of the model is seen as analogous to one experimental trial for a real participant. In an experiment, the experimenter averages over a series of trials to produce the mean performance of a particular real participant. This is repeated for each participant to produce mean group performance. Similarly, a modeler can average over an equivalent number of simulation runs to produce the mean performance of what Hintzman (1986, p. 416) termed a "pseudo-subject." Averaging over a group of pseudo-subjects is thus equivalent to averaging over a group of real participants and permits the same types of analyses.

With this type of modeling, intraindividual variability among pseudo-subjects is the result of random variation in generating the values of the feature vectors and represents randomly fluctuating variation of neural firing patterns accompanying the encoding of each item (see Nesselroade & Salthouse, 2004, for a discussion of whether intraindividual variability should be viewed as noise). In contrast, interindividual variability is the consequence of different starting parameters for each pseudo-subject and is held constant for each pseudo-subject.

Immediate Memory Effects

PSEs and WLEs are defined by a general diminution in performance rather than a qualitative change in results (except where they interact with other factors, e.g., Jones et al., 2007; Neath & Nairne, 1995). We report drops in performance associated with the experimental manipulations by participants and compare these to the predictions derived from running the model across a sample of pseudo-subjects. Following Logie et al. (1996) and Neath, Farley, and Surprenant (2003), effects are presented in terms of the difference between the two conditions divided by performance in the control or standard condition (for further discussion, see Neath & Surprenant, 2007). For example, in what follows the PSE is given by

$$PSE = \frac{D - S}{D}, \tag{8}$$

where D is mean performance in the dissimilar condition and S is mean performance in the similar condition. Similarly, the WLE is given by

$$WLE = \frac{Sh - L}{Sh}, \tag{9}$$

where Sh is mean performance in the short words condition and L is mean performance in the long words condition.

Experiment 1

Experiment 1 compares recall performance of lists of phonologically similar items with recall of lists of dissimilar items. The major difference between this and other published studies is the number of participants.

Method

Participants. A total of 100 Purdue University undergraduates who were native speakers of American English participated in exchange for course credit.

Stimuli. The to-be-remembered items were 64 one-syllable words used by Surprenant, Neath, and LeCompte (1999) to study PSEs. An example similar list is *vote, boat, goat, float, note,* and *coat*. An example dissimilar list is *break, sick, vote, greet, rat,* and *fun*. Each word appeared in both similar and dissimilar lists (e.g., *vote* in the above example). Lists were randomized such that participants saw the same words but in different random orders.

Procedure. Participants were informed that we were interested in how accurately they could remember the order in which a series of words had been presented. Each word was shown in uppercase, center justified, for 1.5 s in the middle of the screen in 20-point Helvetica font. After the final word was shown, six response buttons became active and were labeled with the six words in alphabetical order. The participants were asked to indicate the presentation order by clicking on appropriately labeled buttons using the mouse. For example, if they thought the first word was *break*, they should click on the button labeled *break* first. If they thought the third word was *vote*, they should click on the button labeled *vote* third. Participants received 20 lists, half with dissimilar and half with similar items, and were informed they could take rest breaks at any point. The order of dissimilar and similar trials was randomly determined for each participant. Participants were tested individually, and an experimenter remained in the room to ensure compliance with the instructions.

Results and Discussion

Participants' order reconstruction (calculated as mean proportion of items in the correct order) was more accurate on dissimilar (.75) than similar (.61) lists, the typical result. This pattern was confirmed by analysis of variance (ANOVA), $F(1, 99) = 101.89$, $MSE = 0.052$, $p < .05$. The mean PSE was .169 (lower quartile = .068; median = .178; upper quartile = .273; SD = .195; range = −.833 to .625), and the distribution did not differ from normal (Kolmogorov–Smirnov $Z = 0.815$, *ns*). Seven participants performed more accurately on similar than dissimilar words, and 9 showed no difference.

Simulation results are based on 20 runs per pseudo-subject, half with dissimilar and half with similar items, just as with the real participants. Phonological similarity was modeled by setting a random subset of the features to equal values, just as in previous simulations (e.g., Nairne, 1990; Neath, 2000). A full listing of parameters is shown in the Appendix.

Two types of models were compared. For the "mean" model, all parameters were identical for all pseudo-subjects. Any variability in output is due entirely to the particular random processes within the feature model. In contrast, the "variance" model reflects the observation that different participants vary in their overall level of performance. In the feature model, this is captured by the attention–scaling parameter a. This parameter was initially set to the same starting value for each pseudo-subject but was then increased (or decreased) by an amount determined by sampling randomly from a normal distribution with a mean of zero and a standard deviation of four. The resulting value was then used for all simulation runs for that particular pseudo-subject.[1]

Figure 1 compares the fits of the two models to various aspects of the data.[2] Although both models simulated the main effect (upper panel) and both the proportion PSE and proportion of the sample who showed a reversed effect (lower panel), the mean model shows insufficient variability (the upper panel shows standard deviation, the lower panel shows standard error), whereas the variance model shows adequate variability.

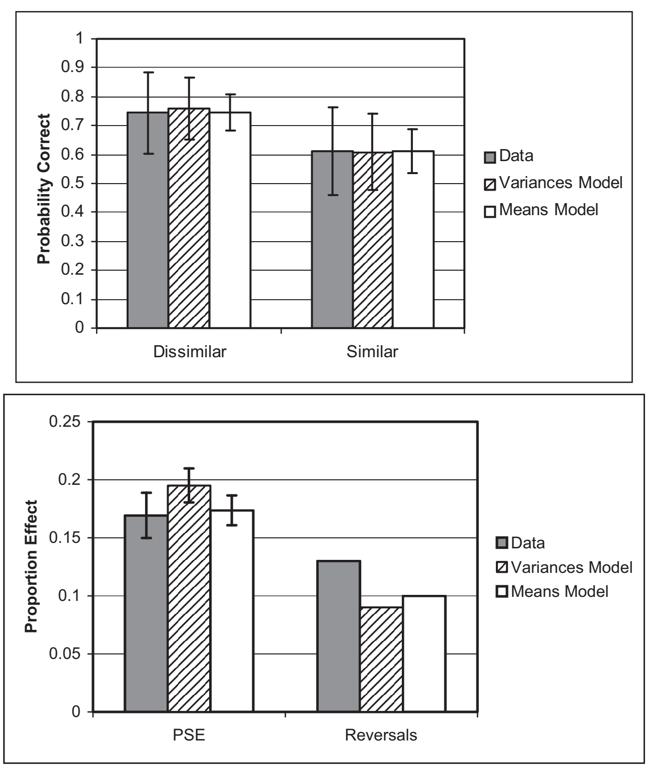


Figure 1. Experiment 1: Fits of a systematic variation of between-participant differences to the experimental data. The upper panel shows the best fits to mean and variance, calculated independently. Error bars show standard deviation, to emphasize this. The lower panel shows the proportion of phonological similarity effects (PSEs) calculated according to Equation 8 and the proportion of the sample showing reversals—superior performance on phonologically similar lists. All data are taken from the simulation runs that produced the results shown in the upper panel. Error bars show standard error.

One way to test the extent to which the feature model is really capturing the data is to examine performance in the upper and lower quartiles (see Figure 2). Participants in the upper quartile in the dissimilar condition show the greatest PSE (.21) and the fewest reversals (.04); those in the lower quartile show the smallest PSE (.11) and the most reversals (.24). The results from the variance model (also shown in Figure 2) show the same pattern.

Finally, it is possible to break the simulation and experimental results down still further and report the data by serial position for each group and each condition. These are shown in Figure 3. Although the ordering of the conditions and the amount of variance are appropriate, the model does not produce sufficiently bowed curves.

The experimental results confirm the basic finding of Logie et al. (1996) that not all apparently "normal" participants show the standard PSE and also provide evidence that variation in the size of the PSE is best described by a normal distribution. This latter finding is what would be expected on the basis of nonstrategic between-participant differences; an account based on two different encoding strategies (phonological and nonphonological) would predict a bimodal distribution. The data go further, however, in indicating that interindividual variation in the size of the PSE is directly attributable to variation in immediate memory capabilities rather than the operation of chance factors. One failing of the model, however, is the insufficient curvature of the serial position curve, particularly in the phonologically similar condition for the top-performing quartile (see Figure 3).

Experiment 2

The second experiment repeats the procedure adopted in Experiment 1, substituting a word length manipulation for a phonological similarity manipulation.

Method

Participants. A total of 100 different Purdue University undergraduates who were native speakers of American English participated in exchange for course credit.

Stimuli. The stimuli were 80 long (three to five syllables) and 80 short (one syllable) words from the study by LaPointe and Engle (1990). Sampling from each set was carried out without replacement.

Procedure. The instructions and procedure were identical to those for Experiment 1 except for the following. Each word was

[1] The parameter a was set to 10 in the mean model and 13.5 in the variance model. The reason for this change was to keep the overall level of performance constant.

[2] As in other simulations with the feature model (e.g., Nairne, 1990; Neath, 1999, 2000; Neath & Nairne, 1995; Neath & Surprenant, 2007), there is no attempt to obtain the best possible quantitative fits. Rather, the goal is to show the correct qualitative pattern of results by changing only those parameters associated with the particular psychological processes implicated. Specifically, we adjusted the model to match the overall performance correct in the "standard" condition (i.e., dissimilar or short) and then examined the rest of the model's predictions.

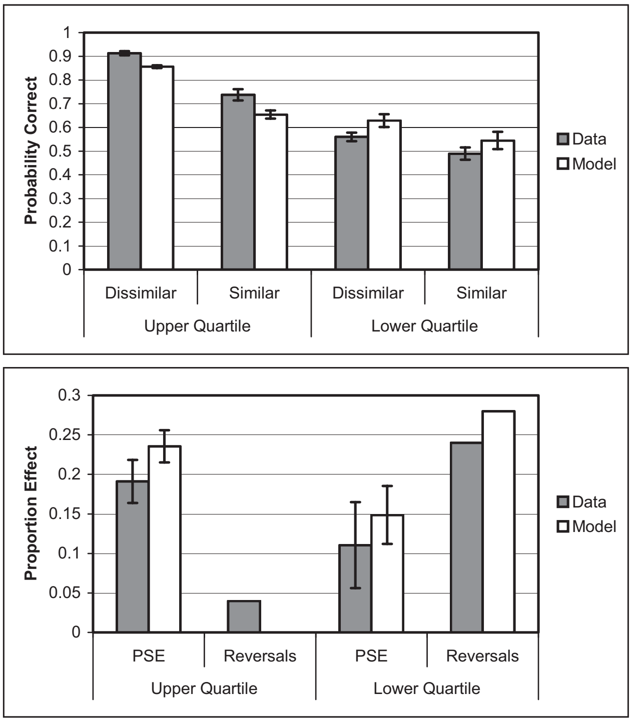


Figure 2. Experiment 1: Data from the experimental sample and the variances model, subdivided into high performance (upper quartile) and low performance (lower quartile) individuals on the dissimilar list condition. The upper panel shows the basic effect for both data and model, and the lower panel shows the phonological similarity effect (PSE) expressed as a proportion and the proportion of individuals among the subgroups showing overall reversal effects. All error bars show standard error.

shown in uppercase for 1 s in the middle of the window on a computer screen in 24-point Helvetica font. The list length was eight, and thus eight response buttons were used. Participants received 40 lists, half with short and half with long items. The order of short and long trials was randomly determined for each participant.

Results and Discussion

Participants correctly recalled in order more short words (.48) than long words (.40), indicating that the word-length manipulation was successful, as confirmed by ANOVA, $F(1, 99) = 103.7$, $MSE = 9.30$, $p < .05$. The mean WLE was .16 (lower quartile = .060; median = .158; upper quartile = .284; $SD = .162$; range = −.30 to .55), and the distribution did not differ from normal (Kolmogorov–Smirnov $Z = 0.428$, *ns*). Thirteen participants showed a reversal, and 3 showed no effect.

Pseudo-subjects were presented with the same number and structure of lists. The length manipulation was simulated as in previous work (e.g., Neath & Nairne, 1995, pp. 432–433). It was assumed that short words consisted of a single segment, long words had seven segments, and the probability of a segment error was .10 for both word lengths. Individual variation was modeled as in Experiment 1. A full listing of parameters is shown in the Appendix.

Figure 4 shows the results of separate simulation runs of sets of pseudo-subjects, using the same procedure as Experiment 1. Again, good fits were obtained for both the mean and variance models, with the latter better reproducing the amount of variability seen in the data.

We again examined how overall performance level affects the appearance of the WLE by once again comparing the upper and lower quartile of the human participants and pseudo-subjects. Figure 5 shows that like the PSE in the previous experiment, there is a trend toward a larger WLE (.19) and fewer reversals (.04) in the better performing group and a smaller WLE (.13) and more reversals (.16) in the lower performing group in both the data and in the predictions of the model. It is important to note that this pattern, predicted by the model, was not obvious a priori.

Serial position curves for the two groups are shown in Figure 6. This figure demonstrates that the model again underpredicts the curvature of the serial position curve. Table 1 summarizes the distribution of PSE and WLE reversals in the experimental and simulation data from the variance model.

The results of this study reveal a similar pattern of effect by performance interaction with word length as appeared for phonological similarity. This relationship also held for the simulation model, despite the different loci for the two effects within the model. In Experiment 2, which made use of longer list lengths, the failure of the model to produce the appropriate degree of serial position curvature was also more noticeable, in particular the lack of primacy over the first two serial positions.

General Discussion

The two experiments reported here confirm that the general pattern of results within Logie et al.'s (1996) span data are replicable when the procedure is changed to one more commonly used in immediate memory experiments. A sizable minority of participants either failed to show or showed reverse PSEs and WLEs. These zero and negative effects were, nonetheless, predicted by a model designed to show detrimental effects of phonological similarity and word length. Furthermore, the model successfully accounts for these data and other measures of variability within the data without any qualitative changes in its operation. Previously, it has been suggested that a diminished PSE or WLE reflects a change in strategy (e.g., lack of rehearsal; Baddeley & Larsen, 2007; Logie et al., 1996) or a move to a different form of storage (Baddeley, 2000). This line of reasoning is also used in the neuropsychological literature to relate memory processes to structure (e.g., Silveri et al., 1998). The simulations reported here demonstrate that a merely quantitative change is sufficient to produce such variation in both effect sizes, with the PSE and WSE being associated with high performance levels. However, two potential criticisms might be leveled at the current study. The first is that the simulations fit only statistical noise and, as such, are both (a) theoretically uninteresting and (b) not unique to the model we describe. The second criticism is that any model can provide a better fit to the data if another parameter is provided and that the current simulations do little more than demonstrate this fact.

Random Variation and Post Hoc Parameter Fitting

Individual differences in the size of the PSE and WLE are normally distributed. A normal distribution of effect size with a similar inci-

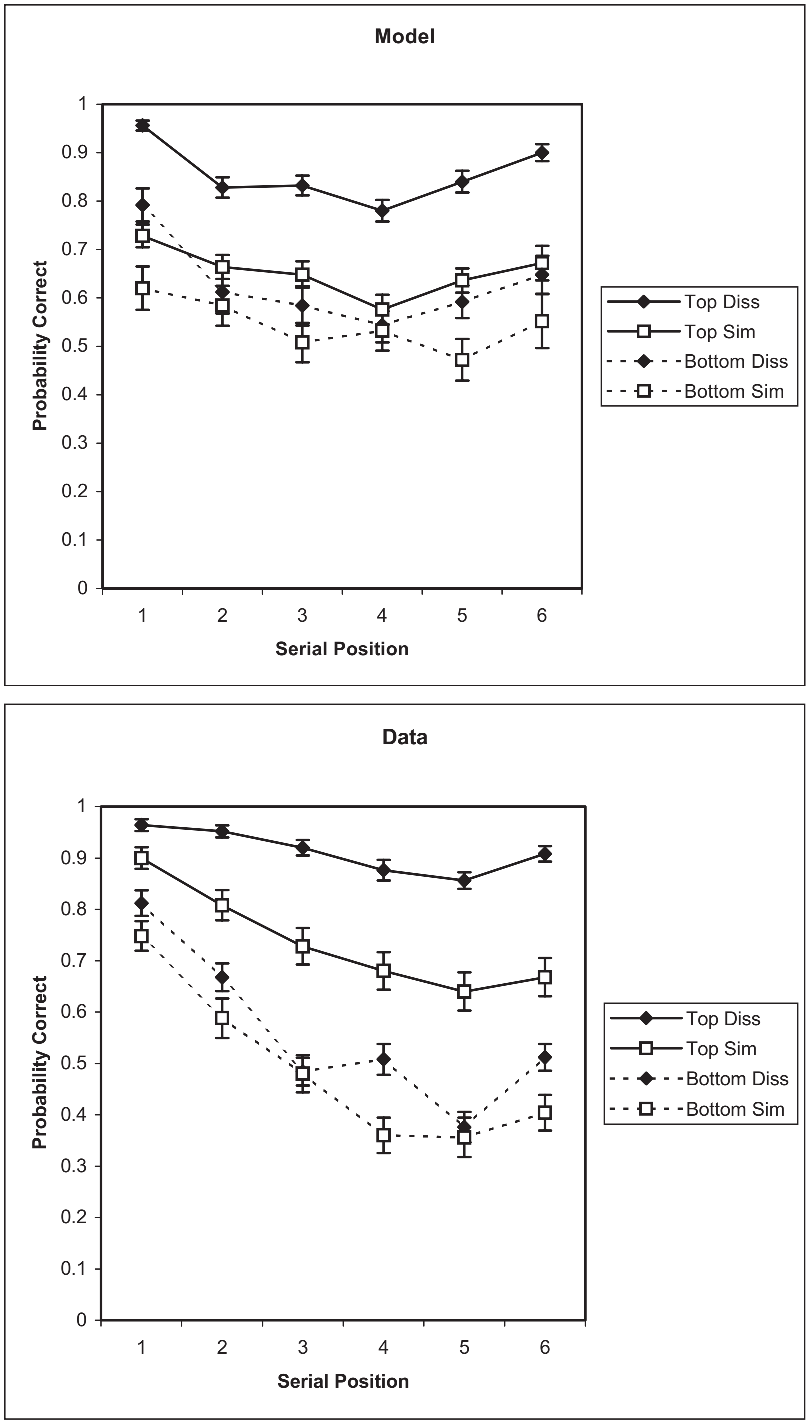
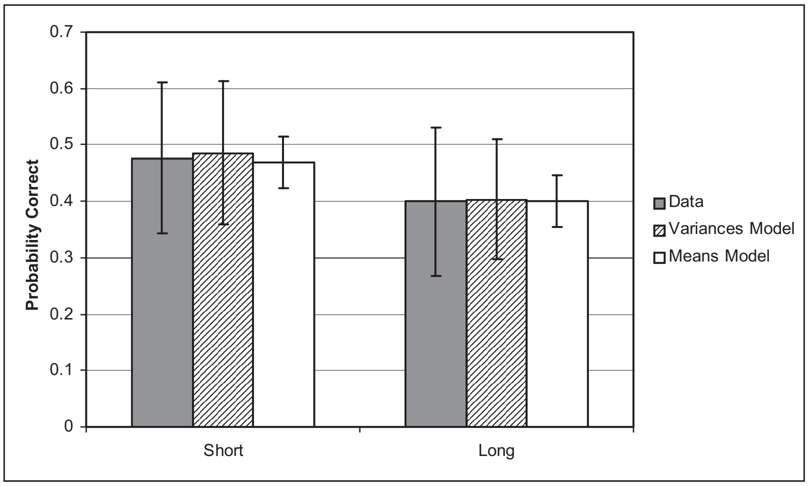


Figure 3. Experiment 1: Results broken down by serial position for the high performance (top) and low performance (bottom) groups in phonologically dissimilar (Diss) or similar (Sim) conditions. Error bars show standard error.

dence of null and reverse effects also results if both interindividual and intraindividual variations are provided by the randomly fluctuating values generated for the feature vectors. This reinforces the central message of this study, that strategy change is not necessary to account for reduction in the PSE and the WLE. In actuality, however, one would expect some reliable difference in performance between Participant A and Participant B, which requires manipulation of a parameter between pseudo-subjects, as reported for the simulations accompanying Experiments 1 and 2, to provide the necessary intraindividual consistency. There is, however, a complication.



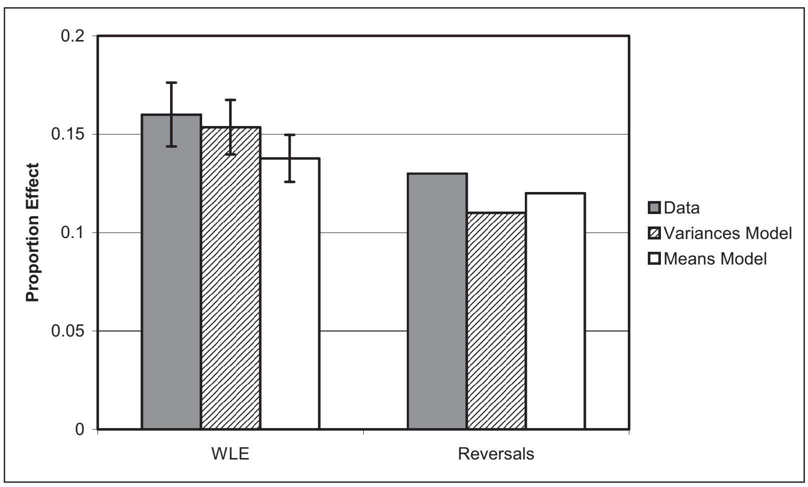


Figure 4. Experiment 2: Fits of a systematic variation of between-participant differences to the experimental data. The upper panel shows the best fits to mean and variance, calculated independently. Error bars show standard deviation, to emphasize this. The lower panel shows the proportion of word length effects (WLEs) calculated according to Equation 8 and the proportion of the sample showing reversals—superior performance on long word lists. All data are taken from the simulation runs that produced the results shown in the upper panel. Error bars show standard error.

As the correlation between components of a difference increases (such as when the component scores are provided by the same individual or dependent on the same parameter within a pseudo-subject), the intraindividual reliability of the difference score itself decreases (Johns, 1981; Peter, Churchill, & Brown, 1993). Ironically, the observation of an association at the interindividual level between memory performance and effect size is the feature that could potentially cast doubt on the reliability of the intraindividual association. The model as reported clearly shows attenuated effects among low ability pseudo-subjects. However, because of the relatively small number of data points per individual in the experimental data it is unrealistic to test, at an intraindividual level, the reliability of the link between memory ability and PSEs and WLEs. Despite this shortcoming, and as noted earlier, a relation between memory ability and memory effects at either intraindividual or interindividual levels militates against the suggestion that reductions in the PSE and the WLE are necessarily consequences of altered strategy—unless that strategy can be shown to be the most effective of any adopted. The feature model provides an alternative, and arguably more parsimonious, mechanism, but further research is clearly necessary to fully test the viability of this alternative mechanism.

The next points to consider are whether the predictions of the feature model are unique to that model, and whether it accounts for the data by something other than post hoc parameter fitting. It is possible that many models could account for variation within an experimental sample. However, no such results have yet been reported, and it is incumbent on those who wish to claim that their models produce the appropriate distributions to demonstrate that interindividual variation is within their capabilities. The more serious point is whether a model of quantitative variation accounts for the systematic relationship between memory ability and the effect sizes in any theoretically satisfying way. The simulations of the feature model show that the PSE and the WLE are expected to

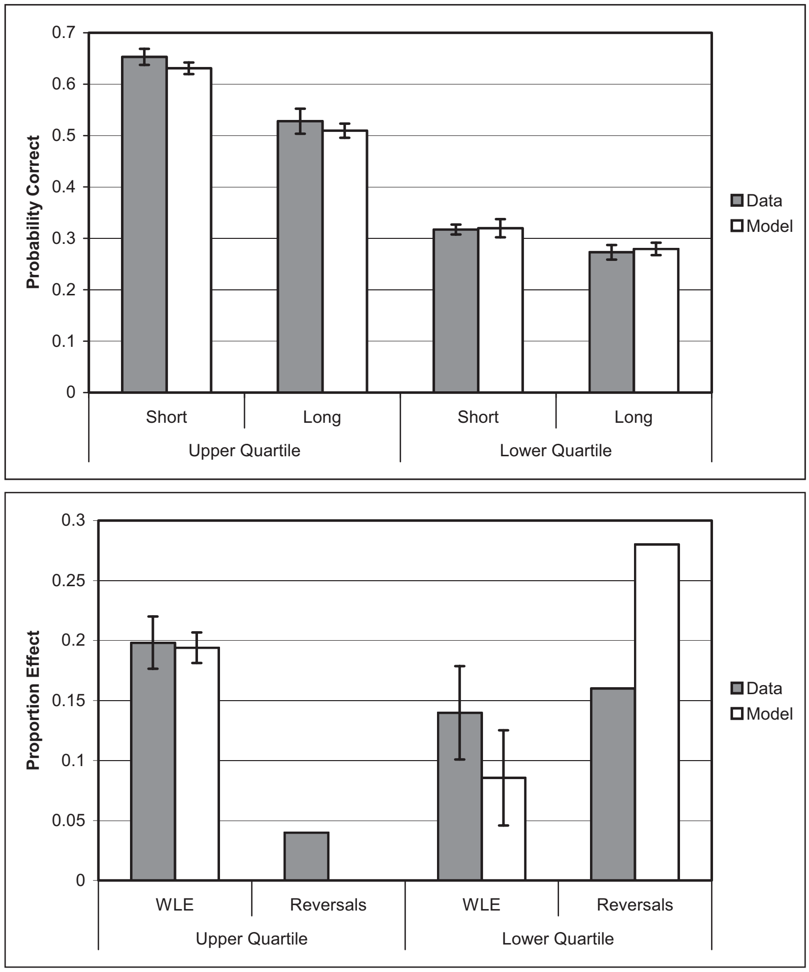


Figure 5. Experiment 2: Data from the experimental sample and the variances model, subdivided into high performance (upper quartile) and low performance (lower quartile) individuals on the short words condition. The upper panel shows the basic effect for both data and model, and the lower panel shows the word length effect (WLE) expressed as a proportion and the proportion of reversals shown among the subgroups. All error bars show standard error.

vary as a function of quantitative changes in participants' abilities, a prediction that was confirmed by examination of the experimental data. But is this a unique feature of the model or a piece of post hoc parameter fitting?

The fit to the mean and variance scores from our participant samples (upper panels of Figures 1 and 5) were produced by a set of parameters appropriate to the performance levels observed within the experimental data set. Crucially, the subsequent splitting of the samples into high- and low-performance quartiles was carried out only after these parameters had been determined and were not free to vary. All simulation results beyond this point (Figures 2, 3, 5, and 6 and the lower panels of Figures 1 and 4) comprise a novel, and genuine, prediction of a positive relationship between memory capability and PSEs and WLEs. The reason for this pattern, which was not obvious prior to simulation, is the specific role of the attention parameter in magnifying the distance score between mismatches and potential matches (Equation 1). The attention parameter was used to simulate interindividual variation because it is the only parameter within the model that varies over all performance levels to a significant degree and hence is the only viable source of systematic interindividual variation.[3] Varying this factor between individuals is a necessary requirement to enable systematic interindividual differences, but the factor was neither new nor free to vary once an appropriate distribution of effect sizes was established. The implication for high- and low-performance subgroups represented a real, and potentially falsifiable, prediction that was, however, confirmed when tested against experimental data. It is not clear that varying a single parameter within other models would, of necessity, produce the same pattern of results, although this can be definitively determined only by extensive simulation studies (by varying the threshold, the primacy gradient, or the level of Gaussian noise within the primacy model; e.g., see Page & Norris, 1998). It is also unclear what parameters

[3] The model used here was the variance model.

might provide a theoretically motivated basis for interindividual differences (as the choice of potential parameters within the primacy model example indicates) or whether it would be necessary to vary different parameters independently to replicate the memory by effect interactions for PSEs and WLEs. In experimental terms, a single quantitative variable within the model we examined necessarily alters the pattern of immediate memory effects observed and is by itself sufficient to cause a qualitative shift in the appearance of PSEs and WLEs. This observation has a number of implications.

Generally, it suggests that purely quantitative variations in some performance factor at either intraindividual or interindividual levels could frequently appear within the experimental data as qualitative shifts in performance characteristics. This reinforces the idea that quantitative computational simulations are a productive method for testing (often implicit) theoretical assumptions that might otherwise go unchallenged or unrecognized (Lewandowsky, 1993). More specifically, the results imply that neither a shift in rehearsal patterns nor the deployment of particular storage buffers (Baddeley & Larsen, 2007) is necessary to explain reductions in the WLE and the PSE among individuals undertaking immediate memory tasks. The feature model predicts both effects using a single PM and no maintenance rehearsal. It is not possible to convincingly show an absence of strategy, so it is incumbent on those who would support the strategy shift idea to show why it is needed to account for differences in effect size. Our data, and that of Logie et al. (1996), show that most, if not all, of the variation in effect sizes can be accounted for directly on the basis of performance characteristics rather than indirectly by performance-based strategy shifts.

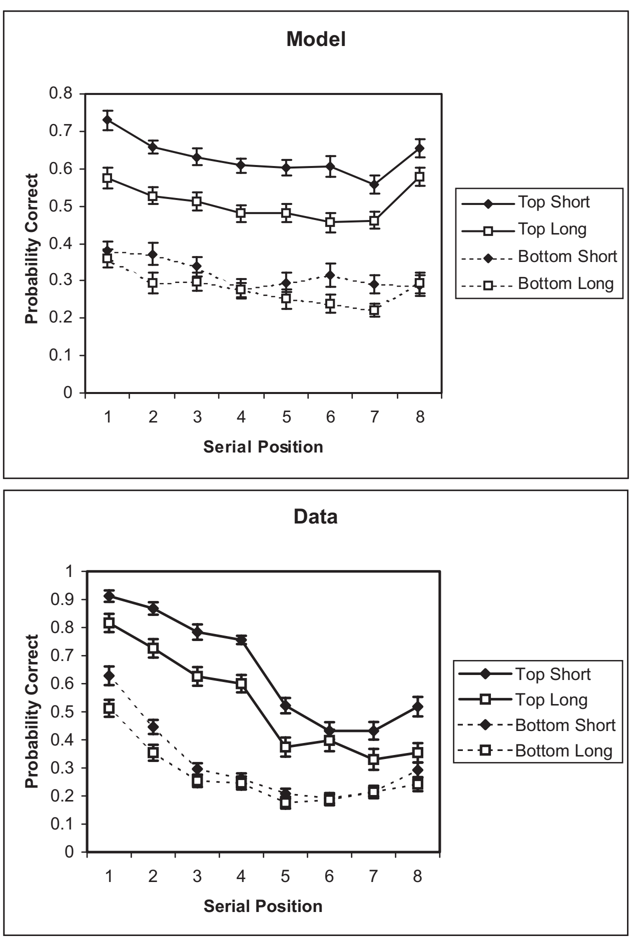


Figure 6. Experiment 2: Results broken down by serial position for the high performance (top) and low performance (bottom) groups in short word (Short) or long word (Long) list conditions. Error bars show standard error.

Table 1
Proportion of the Sample Showing Reversals of the Phonological Similarity Effect (PSE) and Word-Length Effect (WLE) Among the Upper and Lower Quartiles of the Participants and Pseudo-Subjects

	Data		Model	
Quartile	PSE	WLE	PSE	WLE
Upper	0.04	0.04	0	0
Lower	0.24	0.16	0.28	0.28

References

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Larsen, J. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology, 60,* 497–504.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Beaman, C. P. (2006). The relationship between absolute and proportion scores of serial order memory: Simulation predictions and empirical data. *Psychonomic Bulletin & Review, 13,* 92–98.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594–621.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 45*(A), 177–192.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

Cooper, R. P., Schwartz, M. F., Yule, P., & Shallice, T. (2005). The simulation of action disorganisation in complex activities of daily living. *Cognitive Neuropsychology, 22,* 959–1004.

Della Sala, S., & Logie, R. H. (1997). Impairments of methodology and theory in cognitive neuropsychology: A case for rehabilitation? *Neuropsychological Rehabilitation, 7,* 367–385.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society, Series B, 302,* 325–340.

Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 586–594.

Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 98–106.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27,* 443–463.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology, 60,* 505–511.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Lewandowsky, S. (1993). The rewards and hazards of computer simulation. *Psychological Science, 4,* 236–243.

Lewandowsky, S., & Heit, E. (2006). Some targets for memory models. *Journal of Memory and Language, 55,* 441–446.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition, 24,* 305–321.

Lovatt, P., & Avons, S. (2001). Re-evaluating the word-length effect. In J. Andrade (Ed.), *Working memory in perspective* (pp. 199–218). Hove, England: Psychology Press.

Melton, A. W. (1967). Individual differences and theoretical process variables: General comments on the conference. In R. M. Gagné (Ed.), *Learning and individual differences* (pp. 238–252). Columbus, OH: Merrill.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

Neath, I. (1999). Modelling the effects of irrelevant speech on order. *International Journal of Psychology, 34,* 410–418.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7,* 403–423.

Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review, 10,* 430–434.

Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 1269–1278.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2,* 429–441.

Neath, I., & Surprenant, A. M. (2003). *Human memory: An introduction to research, data, and theory* (2nd ed.). Belmont, CA: Wadsworth.

Neath, I., & Surprenant, A. M. (2007). Accounting for age-related differences in working memory using the feature model. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory: Behavioural and neural correlates* (pp. 165–179). Oxford, England: Oxford University Press.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journal of Gerontology: Psychological Sciences, 59*(B), 49–55.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research, 19,* 655–662.

Silveri, M. C., Di Betta, M. M., Filippini, V., Leggio, M. G., & Molinari, M. (1998). Verbal short-term store-rehearsal system and the cerebellum: Evidence from a patient with a right cerebellar lesion. *Brain, 121,* 2175–2187.

Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity and presentation modality. *Memory, 7,* 405–420.

Vallar, G., & Shallice, T. (1990). *Neuropsychological impairments of short-term memory.* Cambridge, England: Cambridge University Press.

Appendix

Parameter Values for Reported Simulations

Feature	Dissimilar	Similar	Short	Long
Modality independent				
No.	20		20	
Range	4		2	
No. guaranteed similar	0	8	0	
Modality dependent				
No.	2		2	
Range	4		2	
No. guaranteed similar	0	1	0	
No. of segments	0		1	7
Probability of segment error	0.1		0.1	
Overwriting probability	1		1	
Attention	10		10.75	
Number of perturbation opportunities	5		3	
θ	0.05		0.05	
Recovery constant	2		2	
Number of recovery attempts	2		2	

Note. Values are shown in the Similar (and Long) column only when they differed from those for the Dissimilar (and Short) condition. For the variance simulations, the attention parameter was set to 13.5 and 11.5 for the acoustic confusion and word-length simulations, respectively, and then increased (or decreased) by an amount determined by sampling from a normal distribution with a mean of 0 and a standard deviation of 4.

Received September 28, 2006
Revision received July 26, 2007
Accepted July 30, 2007 ■