

## **Treating participants (or items) as random vs. fixed effects**

**Dan Mirman**  
**Pyeong Whan Cho**  
**Allison Britt**

### **Abstract**

Participants (or items) are usually treated as random effects, but were treated as fixed effects in our original growth curve analysis article (Mirman et al., 2008). Here we (briefly) explain the difference between fixed and random effects and demonstrate the consequences of treating participants as fixed vs. random effects by re-analyzing data from a previous study (Mirman & Magnuson, 2009). Participants (or items) should be treated as random effects when the primary goal is generalization to a broader population from which they were sampled. However, treating participants as fixed effects may be justified by the properties of the sample (non-random, non-homogenous, or non-normal) or when the primary goal is description of observed data.

## Treating participants (or items) as random vs. fixed effects

When formulating a multi-level polynomial regression model, the researcher must choose whether to treat participants (or items, for a by-items analysis) as fixed or random effects. The traditional logic is that if a factor is interesting in itself and its levels are fixed in the world and reproducible (i.e., experimentally-controlled factors such as word frequency, participant's native language, etc.), they should be considered fixed effects; if the levels correspond to randomly-sampled observational units (e.g., individual participants from some population or words from a set with particular properties), then they should be considered random effects. This is the familiar approach of t-tests and ANOVA's and extended to multi-level regression for VWP data by Barr (2008). However, in our description of growth curve analysis (GCA), we (Mirman et al., 2008) treated both experimentally-controlled factors (e.g., word frequency) and participants as fixed effects.

The critical difference is that when participants are treated as a fixed effect, each participant's fixation proportion curve parameters are estimated independently. When participants are treated as a random effect, each participant's fixation proportion curve parameters are constrained to be random deviations from the population mean curve parameters, with the deviations assumed to conform to a normal distribution with mean equal to 0. This additional constraint means that each individual's parameter estimates from a random participant effects model are weighted averages of the parameter estimates from a fixed participant effects model and the group-level parameter estimates. Put simply, the parameter estimates reflect both the individual participant's data and the whole group data. As a result, they tend to "shrink" toward the population mean. In other words, each participant's individual random effect parameter estimates are influenced by the other participants' data. This shrinkage can have positive and negative consequences. When individual participant estimates are allowed to be fully independent (i.e., treated as fixed effects), they provide better (that is, independent) estimates of differences between individual participants, but the resulting model can overfit the data. Here we compare analysis of a simple semantic competition data set (from Mirman & Magnuson, 2009) to show concretely what is the same and what is different when participants are treated as a fixed or random effect.

### Methods

The data were analyzed using GCA (Mirman et al., 2008) with fourth-order orthogonal polynomials, treating participants as either fixed or random effects. Both models also included intercept, linear, and quadratic<sup>1</sup> random effect terms for participants-by-condition. In all other respects, the analysis followed the standard GCA approach described by Mirman et al. (2008) and used in the original analyses of these data (Mirman & Magnuson, 2009).

### Results and Discussion

Table 1 shows the parameter estimates for the fixed effect of condition (Related vs. Unrelated) from the two models. The parameter estimates for these condition effects were identical for the two versions, which is as we would expect given the balanced within-subject design (i.e., participants were orthogonal to the condition manipulation, regardless of whether they were treated as a fixed or random effect). Because of the additional constraints of treating subjects as random effects, they capture less variance,

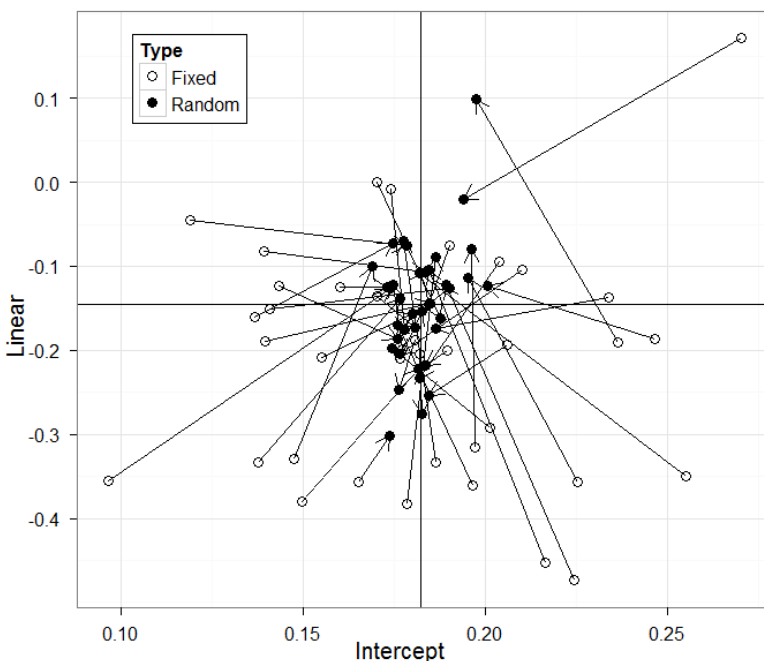
---

<sup>1</sup> This quadratic term was not included by Mirman and Magnuson (2009), but we feel it is important to include it because of the importance of the quadratic term for these competition effects. None of the substantive results depend on the inclusion or exclusion of this term.

thus the standard errors for the condition parameter estimates are larger. The shrinkage effect is illustrated in Figure 1, which shows individual participant intercept and linear slope parameter estimates from the two models. The striking pattern in Figure 1 is the much tighter clustering of individual participant parameter estimates around the population-level fixed effect (indicated by the black vertical and horizontal lines) when the model treated participants as a random effect.

Table 1. Condition effects with participants treated as fixed effect and random effect.

	Participants as fixed effect			Participants as random effect		
	Estimate (SE)	<i>t</i>	<i>p</i> <	Estimate (SE)	<i>t</i>	<i>p</i> <
Intercept	0.070 (0.0087)	8.0	0.00001	0.070 (0.012)	5.8	0.00001
Linear	0.124 (0.029)	4.3	0.0001	0.124 (0.039)	3.2	0.01
Quadratic	-0.111 (0.020)	5.5	0.00001	-0.111 (0.027)	4.1	0.0001
Cubic	-0.041 (0.015)	2.8	0.01	-0.041 (0.016)	2.6	0.01
Quartic	0.066 (0.015)	4.5	0.00001	0.066 (0.016)	4.3	0.0001



**Figure 1.** Shrinkage effect on individual participant intercept and linear term parameter estimates. For each participant, the arrow shows the change in the parameter estimate from a model that treats participants as fixed effects (open circles) to a model that treats participants as random effects (filled circles). The black vertical and horizontal lines indicate population-level fixed effect.

Not surprisingly, because of its greater flexibility, the participants-as-fixed-effects model produced a much better model fit, which is indicated by substantially higher log-likelihood of the model (Participants as fixed effect: LL = 1189; Participants as random effect: LL = 1024). The log-likelihood always increases when independent parameters are added to a model, but we can test whether the additional parameters are justified by the improvement in model fit by evaluating the change in the deviance statistic ( $-2$  times the log-likelihood), which is distributed as chi-square, with degrees of freedom equal to the number of parameters added. In this case, the additional participant fixed effect parameters do significantly improve model fit ( $\chi^2(170) = 329$ ,  $p < 0.0001$ ). This indicates that the participants in the study varied in some way that was not captured by random effects, perhaps language ability or other cognitive skills. However, in many studies (including the study from which these data were drawn), such individual variability constitutes noise because researchers are interested in generalizing from their sample to a larger population. In such cases, participants should be treated as random effects because then the statistical model will correctly instantiate the research question. Note

that in this case, the substantive results were the same whether participants were treated as fixed or random effects, but this will not always be the case.

Researchers deciding whether to treat participants as random or fixed effects need to consider their research goals and their confidence in the homogeneity and normality of the sample population. If the goal is to generalize, then the researcher is essentially forced to assume that the sample is drawn from a homogeneous population and should treat participants as a random effect. However, this form of generalization is not always the goal. For example, neurological case studies<sup>2</sup> inform cognitive theories by showing what must be possible (as in an existence proof) and generating new hypotheses. In such contexts, the goal is to describe the observed data as well as possible and treating participants as fixed effects may be more appropriate. Since participant fixed effect parameters better capture individual differences, they may provide a better approach for studying individual differences (e.g., Mirman et al., 2011; and the individual differences example in Mirman et al., 2008). In such cases, it may be advantageous to acquire independent parameter estimates for the participants by treating them as fixed effects rather than random effects. Finally, for hypothetically homogeneous populations like typical college students, treating participants as random effects may be the better approach; but for clearly non-homogeneous populations like neurological patients (who all have unique clinical and neurological presentations, even if their diagnosis is the same) treating participants as fixed effects may be more appropriate. We note also that model convergence appears to be somewhat more robust for models treating participants as fixed effects rather than as random effects. Thus, if a model with participants as random effects fails to converge, it may be worthwhile to treat them as fixed effects when the alternative is no analysis at all.

In sum, for typical VWP experiments, treating participants (or items) as random effects appropriately reflects the typical assumption that each observational unit is a randomly-drawn sample from a general population to which the researcher hopes to generalize. Treating participants as fixed effects is a legitimate alternative, but should be explicitly justified based on sample properties (e.g., non-random sampling from a non-homogeneous or non-normal distribution) or research goals (e.g., description of present data rather than generalization to a population).

## References

- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. doi:10.1016/j.jml.2007.09.002
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494.
- Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, 37(7), 1026–1039. doi:10.3758/MC.37.7.1026
- Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, 117(2), 53–68. doi:10.1016/j.bandl.2011.01.004
- Teichmann, M., Turc, G., Nogues, M., Ferrieux, S., & Dubois, B. (2012). A mental lexicon without semantics. *Neurology*, 79, 1–2.

---

<sup>2</sup> For a very brief and interesting recent example, see Teichman et al. (2012).