"The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem

Marc Brysbaert

Royal Holloway, University of London

Report written for RTN-LAB¹ Version 1.0, Wedneday, March 28, 2007

Address: Marc Brysbaert Royal Holloway, University of London Department of Psychology Egham TW20 0EX United Kingdom marc.brysbaert@rhul.ac.uk

¹ This report may be distributed freely for educational and research purposes. It does have copyright, though, meaning that you cannot present it as your own work. If you found the report helpful in the analysis of your data, it would be kind to acknowledge so by citing it using the form: Brysbaert, M. (2007). *"The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem (Version 1.0).* Report Royal Holloway, University of London. The report is available on the internet.

1. Introduction

In recent years psycholinguists have been criticized for using suboptimal statistical tests (Baayen, Davidson, & Bates, 2006; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999). In particular, the use of F1 and F2 tests "to generalize over participants and items" has been called into question. At the same time, rumors are spreading about a much better type of analysis few people understand. In this paper I try to translate my (limited) knowledge in a form that is easy to master, because it consists of a series of cookbook recipes. It is the form used increasingly in stats courses and can be defended on the basis that there are different levels of understanding (e.g., knowing how to work with a statistical package and how to interpret the results vs. being able to build one). My discussion is limited to SPSS, not because I am particularly happy with this package, but because it is most widely used.

2. Why does one need to bother about variance between items?

For a beginning researcher it is tempting to limit the statistical analysis of psycholinguistic data to an analysis based on the average per condition per participant. For instance, if 10 participants make a lexical decision to 5 low frequency words and 5 high frequency words, we will calculate the mean of the reaction times (RT) to the correctly identified low frequency words and the mean of the RTs to the correctly identified high frequency words (in addition to the percentage of errors, which will be used as a second variable). Table 1 shows some results we may obtain (empty cells are errors made by the participants).

🛅 Exa	Example data LDT - SPSS Data Editor											
File Ed	File Edit View Data Transform Analyze Graphs Utilities Window Help											
2												
1 : particip	1 : participant 1											
	participant	Low1	Low2	Low3	Low4	Low5	High1	High2	High3	High4	High5	var
1	1	655	847		687	603	652		706	633	593	
2	2	724	954	653	624	613	649	642	505	659	725	
3	3	589	763		688	589	639		638	596	631	
4	4	647	712	769	594		714	566	684	652	545	
5	5	842		698	711	657	598	639	652	681	684	
6	6		863	647	659	688	655	685	701	706	576	
7	7	711	712	589	624	637	689	625		599	703	
8	8	652	914	723	599	725	675	750	692	618		
9	9	483	752	642	602	568	497	504	615	587	605	
10	10	756	811	699	705	718	637	649	587	675	636	
11												

 Table 1 : Example data of a lexical decision experiment containing of 5 low frequency and

 5 high frequency words. Ten participants in total.

When we calculate the mean RTs of the correct trials for the low and the high frequency words, we get Table 2.

	Participant	LowFreq	HighFreq
1	1	698	646
2	2	714	636
3	3	657	626
4	4	681	632
5	5	727	651
6	6	714	665
7	7	655	654
8	8	723	684
9	9	609	562
10	10	738	637

Table 2 : Mean RT of the low frequency and the high frequency words per participant(correct trials only).

To run the analysis, we have to use an ANOVA with a repeated measure. The figures below show how we get there.

ile Ed	it View Da	ita Transfo	rm	Analyze Graphs Utilities	Wi	ndow Help	2		
🗃 🕞	ant		4	Reports Descriptive Statistics Tables	*]			
	Participant	LowFreq	Hij	Compare Means		var	var	var	
1	1	698		General Linear Model		Univariate			
2	2	714		Mixed Models	F	Multivaria	te		
3	3	657		Correlate	•	Reneated	Measures		
4	4	681		Regression				_	
5	5	727		Loglinear		Variance Components			
6	6	714		Classify				_	
7	7	655		Data Reduction					
8	8	723		Cable					
9	9	609		Stale Necessary					
10	10	738		Nonparametric Tests					
11				Time Series					
12				Survival	•				
13				Multiple Response	•				
14				Missing Value Analysis Complex Samples					
15									
16									
17									

🗰 Exa	mple dat	ta LDT F	1 analysi	is - SPSS	i Data Ec	litor			
File Ed	it View Da	ata Transfo	rm Analyze	Graphs	Utilities Wir	ndow Help			
2	8 🔍 🗠	~ % 🕻	M > [F 0	1			
1 : Participant 1									
	Participant	LowFreq	HighFreq	var	var	var	var	var	var
1	1	698	646						
2	2	714	636		C				
3	3	657	626		Repe	ated Me	easures D	efin 🕑	K
4	4	681	632) (Ohio (Subject Franker I	Name I		-
5	5	727	651		within-s	Subject Factor	Name: Ifreq	Define	
6	6	714	665		Number	of Levels:	2	Reset	
7	7	655	654			Itea(2)		Canaal	
8	8	723	684		Add			Cancer	
9	9	609	562		Chang	ie i		Help	
10	10	738	637		Damas				
11					nemov				
12									
13					Measur	e Name:			
14						1			
15					Add				
16					Chang	e			
17									
18					Hemov	/8			
19							1 I		-
20									
21			8		-		1		

	1 24 22		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	and the		1					
articipan		1									
P	articipant	LowFreq	HighFreq	VAL	Vár S	< Var	- Var	C. War	Yar	vár -	147
1	1	698	646								
2	2	714	636								
3	3	657	626								
4	4	681	632								
5	5	727	651	6							
6	6	714	665	= Re	epeated	Measure	5			l l	٢.
7	7	655	654	-				Marking B.	14		7
8	8	723	684	P	aticipant		within-Subjecti	variacies (n	edt	OK	
9	9	609	562				LouFord11			Paste	1
10	10	738	637			لشاشا	HighFred(2)				1
11						1000				rieset	1
12										Cancel	
13										Help	1
-14											
15											
16											
17							Between-Subie	cts Factorist			
18								and a second			
19						1					
20						80 - 10 I					
21							Councillar				
22						an	are remained.				
23											
28						_					
10.77											_



Tests of Within-Subjects Effects

Measure: MEASURE_1	
--------------------	--

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
freq	Sphericity Assumed	13676.450	1	13676.450	35.646	.000	.798
	Greenhouse-Geisser	13676.450	1.000	13676.450	35.646	.000	.798
	Huynh-Feldt	13676.450	1.000	13676.450	35.646	.000	.798
	Lower-bound	13676.450	1.000	13676.450	35.646	.000	.798
Error(freq)	Sphericity Assumed	3453.050	9	383.672			
	Greenhouse-Geisser	3453.050	9.000	383.672			
	Huynh-Feldt	3453.050	9.000	383.672			
	Lower-bound	3453.050	9.000	383.672			

So, on the basis of our ANOVA with one repeated measure, we get a significant effect: F(1,9) = 35.646, MSe = 383.672, p < .001, Eta Squared = .798⁻². The effect is extraordinarily strong because no participant has a lower mean RT for the low frequency words than for the high frequency words. This is strong evidence that high frequency words are easier to process than low frequency words, isn't it?



Figure 1 : Mean lexical decision time per word: the first five words are the low frequency words; the final five are the high frequency words.

Figure 1 shows another part of the story, however. This figure displays the mean RT per word stimulus. Now, the evidence suddenly looks less impressive: Nearly all the difference between the high and the low frequency words is due to the long RTs for word Low2 (see also Table 1). If we took another sample of words that does not include word Low2, would we still find a frequency effect?

The discrepancy between Table 2 and Figure 1 is what Clark (1973) called "the languageas-fixed-effect fallacy". If we limit our statistical analysis to the analysis reported above, we assume that there is no variability in the words we have chosen, or that our sample exhausts all possible words we could have selected. Given that this rarely is the case, Clarke argued that in our statistical analyses we have to take into account the variability due to the items in addition to the variability due to the stimulus items. Although his analysis is not that difficult to understand, it requires the reader to know something about the difference between fixed and random effects in ANOVAs and about how to calculate Mean Square terms and F-values. In addition, the analysis Clarke proposed (a quasi-F ratio or F') only works when there are no missing data (i.e., when the participants make no errors or when the missing RTs are estimated).

² Eta squared is an index of the effect size. You get it when you click on **Options** and **Estimates of effect** size. The eta squared has a similar meaning as R^2 (how much of the variance is due to the effect). In psychology, most values of eta squared will be around .09 (i.e. r = .30, medium effect size). The high value in the present example gives away that it was constructed by hand.

Luckily, Clarke (1973) also included an easier way around the problem (although Raaijmakers claims this has been one of the big mistakes in psycholinguistic research, because psycholinguists used the shortcut in the wrong way).

The solution Clarke proposed, was to do an F2 analysis in addition to the F1 analysis and to calculate minF'. F1 is the analysis we have discussed above (Table 2). It consists of an ANOVA on the mean values per participant per condition. There can be as many independent variables (IVs) as one likes (although in reality, it is strongly recommended not to have more than 2; higher-order interactions are a nightmare to interpret and usually are unstable; i.e., the exact same pattern is not obtained in a replication of the study, even when the interaction is significant again; in addition, very few researchers have a priori hypotheses about more than two IVs).

	Frequency	RT
1	low	673
2	low	814
3	low	678
4	low	649
5	low	644
6	high	641
7	high	633
8	high	642
9	high	641
10	high	633

Table 3: Mean RT of the 5 low frequency words and the 5 high frequency words.

Table 3 shows the starting point of the F2 analysis, the analysis over items. For this analysis, the researcher calculates the mean RT per word. Because in the present example the words belonging to the high frequency condition and the words belonging to the low frequency condition are different words, the IV will be a between-items variable. These are the steps of the analysis:

File Edit View Data Transform Analyze Graphs Utilities Window Help Image: Imag	🗰 Exa	mple da	ta LDT F2 a	analysis - SPSS Data Editor	
Frequency Reports 1: Frequency Particle 1: Frequency Reports 1 Compare Means Var 2 low 684 3 low 678 4 low 649 5 low 644 7 high 633 8 high 641 10 high 641 10 high 641 11 Classify Data Reduction Scale Nonparametric Tests Immed Multiple Response 11 Multiple Response 12 Multiple Response 13 Multiple Response 14 Complex Samples	File Ed	it View D	ata Transform	Analyze Graphs Utilities Window Help	
Frequency RT Compare Means Var Var Var 1 low 673 General Linear Model Univariate Multivariate 3 low 678 Correlate Repeated Measures Multivariate 4 low 649 Regression Variance Components Variance Components 6 high 641 Classify Variance Components Variance Components 9 high 642 Data Reduction Scale Variance Components 10 high 641 Norparametric Tests Time Series Variance Components 11 Scale Multiple Response Multiple Response Variance Complex Samples 14 Complex Samples Complex Samples Image: Complex Samples Image: Complex Samples	Image: Second			Reports Descriptive Statistics Tables	
1 Iow 673 General Linear Model Univariate 2 Iow 678 Mixed Models Multivariate 4 Iow 678 Correlate Repression 5 Iow 644 Correlate Repression 6 high 641 Casify Variance Components 7 high 641 Casify Data Reduction 8 high 641 Scale Scale 10 high 633 Time Series Survival 11 Survival Multiple Response Multiple Response 14 Corplex Samples Survival Survival 16 Corplex Samples Survival Survival		Frequency	RT	Compare Means	
2 low 814 3 low 678 4 low 649 5 low 644 6 high 641 7 high 633 8 high 642 9 high 641 10 high 633 11 Scale 12 Survival 13 Multiple Response 14 Mising Value Analysis 15 Complex Samples	1	low	673	General Linear Model 🔸 Univariate	
3 low 678 4 low 649 5 low 644 6 high 644 7 high 633 8 high 642 9 high 644 10 high 643 11 Scale 12 Survival 13 Multiple Response 14 Mising Value Analysis 15 Correlate 16 Correlate	2	low	814	Mixed Models Multivariate	
4 low 649 Regression Variance Components 6 high 641 Loginear Variance Components 7 high 633 Classify Variance Components 9 high 641 Scale Variance Components 10 high 641 Scale Variance Components 11 Scale Variance Components Variance Components 11 Scale Variance Components Variance Components 11 Scale Variance Components Variance Components 12 Scale Variance Components Variance Components 13 Multiple Response Variance Complex Samples Variance Components 14 Missing Value Analysis Complex Samples Variance Complex Samples 17 Variance Complex Samples Variance Complex Samples Variance Complex Samples	3	low	678	Correlate Repeated Measures	
S low 644 Loginear Variance Components 6 high 641 Classify 7 high 633 Classify 8 high 642 Data Reduction 9 high 641 Scale 10 high 633 Time Series 11 Survival 13 Multiple Response 14 Complex Samples 16 Complex Samples 17	4	low	649	Regression	
6 high 641 Classify 7 high 633 Classify 8 high 642 Scale 9 high 641 Scale 10 high 633 Nonparametric Tests 11 Survival Time Series 13 Multiple Response Missing Value Analysis 15 Complex Samples 16 17	5	low	644	Variance Components	
7 high 633 Classify - 8 high 642 Data Reduction - 9 high 641 Scale - 10 high 633 Nonparametric Tests - 11 Time Series - - 12 Survival - - 13 Multiple Response - - 14 Missing Value Analysis - - 15 Complex Samples - - 17 - - - -	6	high	641	Claceify	
8 high 642 Data Reduction 9 high 641 Scale 10 high 633 Nonparametric Tests 11 533 Time Series 12 Survival 13 Multiple Response 14 Missing Value Analysis 15 Complex Samples 17 18	7	high	633	Data Reduction	
9 high 641 Scale 10 high 633 Nonparametric Tests	8	high	642		
10 high 633 Nonparametric Lests Image: Comparametric Lests Image: ComparametricLests Image	9	high	641	Scale	
11 Time Series Image: Series 12 Survival Image: Series 13 Multiple Response Image: Series 14 Missing Value Analysis Image: Series 16 Complex Samples Image: Series 17 Image: Series Image: Series 18 Image: Series Image: Series	10	high	633	Nonparametric l'ests	
12 Survival Image: Survival 13 Multiple Response Image: Survival 14 Missing Value Analysis Image: Survival 15 Complex Samples Image: Survival 16 Image: Survival Image: Survival 17 Image: Survival Image: Survival 18 Image: Survival Image: Survival	11			Time Series	
13 Multiple Response Image: Mail and the second se	12			Survival •	
14 Missing Value Analysis 15 Complex Samples 16 17 18	13			Multiple Response 🔸	
Complex Samples Image: Complex Samples 16 Image: Complex Samples Image: Complex Samples 17 Image: Complex Samples Image: Complex Samples 18 Image: Complex Samples Image: Complex Samples	14			Missing Value Analysis	
	15		1	Complex Samples 🕨	
17 18 10	16				
18	17				
10	18				
12	19				





Tests of Between-Subjects Effects

Dependent Variable: RT								
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared		
Corrected Model	7182.400(a)	1	7182.400	2.920	.126	.267		
Intercept	4419590.400	1	4419590.400	1796.837	.000	.996		
Frequency	7182.400	1	7182.400	2.920	.126	.267		
Error	19677.200	8	2459.650					
Total	4446450.000	10						
Corrected Total	26859.600	9						

a R Squared = .267 (Adjusted R Squared = .176)

In the F2 analysis we see that the effect of word frequency is not significant (F2(1,8) = 2.92, MSe = 2460, p = .126, eta squared = .267). In the psycholinguistic community, this "means" that on the basis of the present data we cannot assume that the finding generalizes to other stimuli (notice that it is a null-effect, so the researcher is not allowed to conclude that the effect is 'absent'; the power of the experiment is way too low for that).

Clarke (1973) himself did not pay too much attention to the particular value of F2 (rightly so) and only calculated it because it allowed him to obtain a reasonably good estimate of an F value that would *generalize at the same time across participants and items*, which he called the **minF**'. The minF' value is calculated as follows:

$$\min F'(i, j) = \frac{F1 * F2}{F1 + F2}$$

 $i = df_1 of F_1 = df_1 of F_2$ (df_1 of F1 has to be the same as df_1 of F2)

$$j = \frac{(F1 + F2)^2}{\left(\frac{F1^2}{df_2 - of - F2} + \frac{F2^2}{df_2 - of - F1}\right)}$$

Applied to our example, this gives

i = 1

$$j = \frac{(35.646 + 2.920)^2}{\left(\frac{35.646^2}{8} + \frac{2.920^2}{9}\right)} = 9.3 \approx 9$$

$$\min F'(1,9) = \frac{35.646 * 2.920}{35.646 + 2.920} = 2.699$$

To find the p-value associated with minF', you can use the built-in Excel function [FDIST(2.699,1,9) = .135] or use a ready-made applet on the internet (see http://www.pallier.org//ressources/MinF/compminf.htm or http://www.pallier.org//ressources/MinF/compminf.htm or http://www.pallier.org//ressources/MinF/compminf.htm or

The minF' test informs us that we are not allowed on the basis of the data in Table 1 to argue for a reliable frequency effect that generalizes both across participants and stimuli, which was Clarke's message.

3. Getting overly excited about F2

In the years after Clark (1973) the importance of an F2 items analysis became generally accepted in psycholinguistics, but gradually psycholinguists forgot about minF' (Raaijmakers et al., 1999). Part of the reason for this was that psycholinguists were not aware of the fact that a significant F1 and a significant F2 do not suffice to get a significant minF' (what Raaijmakers et al. called the "F1 x F2 fallacy"). In addition, there were good reasons to expect that minF' would be a conservative test (i.e., more difficult to get significance with it), although subsequent simulations showed that this problem was less severe than feared at the onset.

Anyway, gradually psycholinguists moved away from minF' and limited their analyses to F1 (to check whether the findings could be generalized across participants) and F2 (to check whether the findings could be generalized across stimulus materials). In addition, psycholinguists became more and more 'sophisticated' in their use of F1 and F2. Looking at Table 3, we see that the F2 analysis in our example is one of the least powerful tests one could imagine: Because word frequency is a between-items variable, all noise due to the individual words is added to the error term and one needs large numbers of observations in the different conditions to find a significant F2. This is in particular a problem when pairs of words have been assembled that differ on one particular variable (e.g., age of acquisition, AoA) and are matched on a list of other variables (frequency, word length, number of orthographic neighbors, ...). Because of the control variables, large differences between the word pairs are expected (otherwise one would not need to match the stimuli on these variables) and this variance should be partialed out before we start the F2 analysis. One solution is to use a repeated measures design for the F2 analysis as well. In this analysis the pairs of stimuli are considered as observations from the same 'entity' or 'block' (analog to the 'participant' in a repeated measures F1 analysis). Table 4 shows how the data of such an F2 design would look like for an experiment in which 10 pairs of words have been selected that differ in AoA (one word is acquired early in life, e.g., daffodil, the other word is acquired late in life, e.g., participant) and both are matched on a series of other measures (frequency, ...).

🗰 Exai	Example data AoA matched pairs F2 analysis - SPSS Data Editor									
File Edi	t View Dat	a Transform A	analyze Graphs	Utilities	Window He	elp				
	696 1 1 1 1 1 1 1 1 1 1									
10:	10:									
	Word_pair	Early_acquired	Late_acquired	var	var	var	var			
1	1	773	825							
2	2	814	856							
3	3	678	745							
4	4	542	624							
5	5	644	659							
6	6	683	705							
7	7	498	615							
8	8	563	672							
9	9	641	648							
10	10	699	753							
11										
12										
13										
14										

For these data, the F2 with repeated measures is F2(1,9) = 22.647, MSe = 710, p = .001, eta squared = .716), whereas with a between-items analysis it would be F2(1,18) = 1.922, MSe = 8366, p = .183, eta squared = .096). The reason for the lack of power of the between-items analysis becomes clear when you compare the mean squares of error of both tests (8366 vs. 710). In the repeated-items analysis, a lot of the variability between the stimuli is partialed out as variability between the blocks, due to variation in the control variables, whereas this variance is included in the error term of the between-items test, making it very hard to find a significant F2 (just like one needs at least 128 participants to look for a medium size effect in a between-participants F1 analysis with 1 IV and 2 conditions).

Another way to 'improve' the F2 analysis is to include a Latin-square variable (Pollatsek & Well, 1995). A technique psycholinguists often use, is to counterbalance their stimuli over participants. Imagine, for instance, that you want to investigate semantic priming. To do so, you search for target words with related and unrelated primes (e.g., using the Edinburgh Thesaurus, <u>http://www.eat.rl.ac.uk/</u> or Nelson's Florida norms, http://w3.usf.edu/FreeAssociation/). These are some of the words you may come up with:

Target	Related prime	Unrelated prime
bread boy nurse	butter girl doctor	buffer curl danger
cat	dog	day

•••

Because you do not want to present your target words twice to the same participant, half of the participants see bread preceded by butter and the other half sees bread preceded by buffer, and so on. So, you will make two stimulus list:

List 1	List 2
butter-bread curl-boy doctor-nurse day-cat	buffer-bread girl-boy danger-nurse dog-cat

Half of the participants will get list 1 and half list 2. Now, a typical problem in such a design is what to do with a slow or a fast participant. Table 4 illustrates what can happen:

🗰 data	a latin sq	uare de	sign sem	iantic p	riming -	SPSS Dat	ta Editor	٢		
File Ed	it View Da	ata Transfo	rm Analyze	e Graphs	Utilities Wir	ndow Help				
	a 🔍 🗠	a 1			5 30					
10 : cat_u	nrelated				and a second second					
	participant	bread_relat ed	bread_unrel ated	boy_related	boy_unrelat ed	nurse_relat ed	nurse_unrel ated	cat_related	cat_unrelat ed	var
1	1	624			694	588			658	
2	2		648	684			675	602		
3	3	1024			1124	968			1214	
4	4	1	655	625		2	645	668		
5	5	745	2	2	684	593	2	2	695	
6	6		698	652			689	656		
7	7	635			674	654			653	
8	8		704	705			695	687		
9	9	674			639	655			708	
10	10		657	658			597	604		
11					1					
12										
13										

The important person here is participant 3, who is considerably slower than everyone else. Because of the Latin-square design, this person will add extra RT to the related condition for the stimuli *butter-bread* and *doctor-nurse*; similarly s/he will add extra RT to the unrelated condition for the stimuli *curl-boy* and *day-cat*. This will show in the data that are entered in the F2 analysis, as can be seen below:

🖩 Untitled - SPSS Data Editor											
File Ed	it View Dat	a Transform	Analyze Graphs	s Utilities \							
B B C C C C C C C C C C											
14 :			<i>.</i>								
	Target_word	related_prime	unrelated_prime	var							
1	bread	740	672								
2	boy	665	763								
3	nurse	692	660								
4	cat	643	787								
5											
6											
7											
8											

For the target stimuli *boy* and *cat*, we find a huge effect in the expected direction, whereas for the stimuli *bread* and *nurse*, we find a small effect in the opposite direction, even though nearly all the individual participants showed the predicted semantic priming effect. If we do the calculations, we find F2(1,3) = .489, MSe = 5158, p = .535, eta squared = .140. Needless to say, such a low F2 value will also result in a low minF'.

One way to increase the power of this design is to add a Latin-square variable to the design. The words *bread* and *nurse* were seen in the related condition by one group of 5 participants, and in the unrelated condition by another group of 5 participants. And vice versa for the words *boy* and *cat*. Therefore, what we can do to get rid of the difference in average RTs between the groups, is to add the following between-items variable:

🖩 Untitled - SPSS Data Editor										
File Edit View Data Transform Analyze Graphs Utilities Window Help										
2										
4 : Latin_S	quare_Group	2								
	Target_word	related_prime	unrelated_prime	Latin Square Group	var					
1	bread	740	672	1						
2	boy	665	763	2						
3	nurse	692	660	1						
4	cat	643	787	2						
5										
6										
7										

Tests of Within-Subjects Effects

Measure: MEASURE_1 Partial Type III Sum Mean Eta of Squares df F Source Square Sig. Squared factor1 Sphericity Assumed 2520.500 1 2520.500 5.910 .136 .747 Greenhouse-Geisser 2520.500 1.000 2520.500 5.910 .136 .747 Huynh-Feldt 2520.500 1.000 5.910 .136 .747 2520.500 Lower-bound 1.000 .747 2520.500 2520.500 5.910 .136 factor1 * Sphericity Assumed 14620.500 34.280 1 14620.500 .028 .945 LS_group Greenhouse-Geisser 14620.500 1.000 14620.500 34.280 .028 .945 Huynh-Feldt 14620.500 1.000 14620.500 34.280 .028 .945 Lower-bound 14620.500 1.000 14620.500 34.280 .028 .945 Error(factor1) Sphericity Assumed 853.000 2 426.500 Greenhouse-Geisser 853.000 2.000 426.500 Huynh-Feldt 853.000 2.000 426.500 Lower-bound 2.000 853.000 426.500

Although the small number of stimuli in our example does not allow us to reach significance, the F2 test now looks much more convincing (F2(1,2) = 5.910, MSe = 426.5, p = .136). A look at the ANOVA table shows that a lot of the noise in the F2 analysis caused by the slow participant 3 has been captured by the interaction effect between semantic priming and Latin-square group. In the same way, unintended variation

between the stimuli that make up list 1 and list 2 can be partialed out by including a Latin-Square variable in the F1 analysis. Another way to get rid of unintended variation due to slow participants, is to use the z-scores per participant (i.e., the (RTs – $M_{participant})/sd_{participant}$, a technique used by Balota and Besner).

4. Being put down again

Just when psycholinguists thought they were getting savvy enough to run proper analyses, they were attacked anew. First, there was Raaijmakers' comment that a significant F1 and a significant F2 were not enough to generalize across participants and stimuli. This prompted JML to require all its authors to report minF' in addition to F1 and F2. As a kind of consolation, Raaijmakers et al. (1999) added that an F2 analysis is not always required and in some cases can even lead to a needless loss of power. Ironically, by doing so Raaijmakers et al. repeated Clark's (1973) mistake, because ever since I've seen more references to Raaijmakers et al. by authors claiming that their non-significant F2 analysis is of no real concern than by authors arguing why they believe minF' is more important than separate F1 and F2 analyses.

At the same time, Baayen started to launch the claim that the minF' analysis as a combination of F1 and F2 is needlessly complicated and should be replaced by mixed-effects (or multi-level) modeling (Baayen, 2007; Baayen et al., 2006). Unfortunately, Baayen's language is so specialized that it took me a few months and the help of others to realize what he was talking about. In particular, I've been able to make headway by comparing Baayen et al. (2006) with Locker, Hoffman, and Bovaird (2007) and by trying to understand what Van den Noortgate and Onghena (2006) were doing. Below you find my current understanding of these techniques. It may be wrong in a number of details (in which case I would appreciate your feedback), but at least it looks pretty convincing to me (at the moment). Here we go.

5. Jumping a few levels higher

Just like an ANOVA at its basis is nothing else than a multiple regression, so you can approach the problem of random participants and random stimuli as a regression problem. You try to predict an observed RT as the end result of (i) a participant, (ii) a stimulus, and (iii) the contribution of one (or more) IVs. So, what you try to do is to see whether your manipulation is explaining anything more than what could be predicted on the basis of the participants and the stimuli. The only real thing you need is an algorithm that goes beyond simple linear regression. Turns out that SPSS has such an algorithm! (At least from version 11 on). It is called MIXED. I will go through the procedure on the basis of Table 1 (LDT to high and low frequency words). The first thing to do is rewrite everything as you would for a multiple regression analysis. So, you have three predictor variables: participant, stimulus word, and frequency condition (the latter is recoded as -.5 for a low frequency word +.5 for a high frequency item; by using this code, you can easily interpret the regression weight). So, this gives the following input file:

🗰 Exam	xample data LDT long version for multilevel - SPSS Data Editor													🗙			
File Edit	View Data	a Transform	Analyze	Graphs U	tilities Wind	dow Help											
🗁 🖬 🖨	1 🖳 🖻		# ¶ <u>*</u> [• • •												
T. Patticipant		1		1					_								
P	articipant	Word Fre	equency	RI	var	var	var	var	var	var	var	var	var	var	var	var	var 🔺
	1		5 E	000					_								
2	1	2	0 E	647													
4	1		0	603													
	1	6	5	652					_								
6	1	8	.5	706													
7	1	9	.5	633													
8	1	10	.0	593					_								
9	2	1	- 5	724													
10	2	2	5	954													
11	2	3	5	653													
12	2	4	5	624													
13	2	5	5	613													
14	2	6	.5	649													
15	2	7	.5	642													
16	2	8	.5	505													
17	2	9	.5	659													
18	2	10	.5	725													
19	3	1	5	589													
20	3	2	5	763													
21	3	4	5	688													
22	3	5	5	589													
23	3	6	.5	639													
24	3	8	.5	638													
25	3	9	.5	596													
26	3	10	.5	631													
27	4	1	5	647													
28	4	2	5	712													
29	4	3	5	769												L	
30	4	4	5	594													
31	4	6	.5	/14					_								
32	4	/	.5	566					_								
33	4	8	.5	664													
34	4	9	.5	652													L
35	4 5	10	.5	545													
30	5	1	5	042					_	_							
37	5	3	5	711													
30	5	4	5	667													
40	5	2	5 E	007													
40	5	7	.0	639				-									<u> </u>
41	5	, 8	.5	652													
42	5	9	.5	681					-	-					-		
44	5	10	.5	684									-	L			_
▲ ► \Dat	ta View (\	√ariable View	1						4	SPSS Proc	essor is rear	tv					
and the second		a	Teller	Con the second	1 million	95		a da a		14 0-10	Si cala			50	Norton		18 an.ar
star	9	* Loo 🔨 🤍		Mult	Tine	- Un 3 5		4u0	A MICL	😈 Data	B cac	van	van	FR		N 2 1 2 1 2 1	18 CC:31

The nice thing about this input is that it makes no great deal if there are a few missing observations. You just skip the line (e.g., word 3 for participant 1). The regression method is reasonably robust against empty cells (at least that's what I've read). Then we have to enter our model. Here it is a bit tricky because you must enter the syntax editor. You do this as follows:

🖬 Example data LDT - 🗄	SPSS	Dat	a Editor			
File Edit View Data Transfor	m Ana	alyze	Graphs Ut	ilities Wind	low Help	
New	•	Da	ata	1 ma		
Open	×.	S	/ntax			
Open Database	•	0	utput			
Read Text Data		Dr	aft Output	Low4	Low5	High1
		S	ript	687	603	652
Save Ci	ri+S	P**	000	624	613	649
Save As		63		688	589	639
Mark File Read Only		12	769	594		714
Display Data File Information			698	711	657	598
Casho Data		63	647	659	688	655
Cacille Data	S.L.	12	589	624	637	689
Stop Processor Ci	(SHEV)	14	723	599	725	675
Switch Server		52	642	602	568	497
Print Preview		11	699	705	718	637
Print Ci	trl+P					
Recently Used Data	÷	\vdash				
Recently Used Files	•					
Exit						
17						
18						

This opens a syntax file. Another, more easy way to open a syntax file is to open a ready made file (or to click on it in windows explorer). Then everything opens automatically in SPSS. In the syntax file you write the following:

example data LDT long version for multilevel - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : Paticipant 1
Participant Word Frequency RT var var var var var var var var
1 🔤 Example LDT frequency multilevel - SPSS Syntax Editor
2 File Edit View Data Transform Analyze Graphs Utilities Run Window Help
5 GET FILE = 'D'\Documents and Sattings\Marc\My Documents\avan\Multilaval modeling\avamnla data LDT long varsion for multilaval cov
7 MIXED rt BY Participant Word Frequency
a //METHOD = REML
10 /PRINT = SOLUTION TESTCOV
11 // ANDOM = INTERCEPT I SUBJECT/Word) COVITYEC(N)
12 /EMMEANS TABLES (Frequency).
13
15
17
18
19
23
24 SPSS Processor is ready
25 3 9 .5 596

First, you have to indicate where the computer can find your data file. Then, you indicate what the dependent variable is of your MIXED program (RT) and which predictor variables (Participant, Word, Frequency). Participant and Word are random variables (i.e., a random sample from the population). Frequency is a fixed effect (you are

interested in these two levels). Basically this is all you have to do. You indicate that each participant and each stimulus word can have a different intercept value (i.e. need more or less time to process) and in addition you want to see whether frequency adds enough weight to be significant. The /EMMEANS command gives you the maximum likelihood estimator of the condition means. Once you've entered everything (do not forget the full stops!) you click on RUN. If everything goes well, this is what you should get (among other garbage):

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	11.747	1284.077	.000
Frequency	1	7.988	2.860	.129

Type III Tests of Fixed Effects(a)

a Dependent Variable: RT.

Estimates of Fixed Effects(b)

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval		
						Lower Bound	Upper Bound	
Intercept	637.8481631	24.4844447	10.318	26.051	.000	583.5204508	692.1758754	
[Frequency=5]	54.1145501	31.9990880	7.988	1.691	.129	-19.6941046	127.9232048	
[Frequency=.5]	0(a)	0						

a This parameter is set to zero because it is redundant.

b Dependent Variable: RT.

Frequency(a)

				95% Confidence Interval				
Frequency	Mean	Std. Error	df	Lower Bound	Upper Bound			
5	691.963	24.517	10.374	637.602	746.324			
.5	637.848	24.484	10.318	583.520	692.176			

a Dependent Variable: RT.

The F-value is given in the first table. The second table contains the t-values of the planned comparisons. The F-value is:

F(1,7.988) = 2.860, p = .129

For the sake of comparison, this was the minF' value we obtained:

 $\min F'(1,9) = 2.699, p = .135.$

Not bad if you look at the ease with which you can do this analysis!!! Baayen et al. (2006) have done quite some simulations with this technique (albeit on an R version of theirs, which gives the same results) and they claim that it is safe (i.e., does not result in spurious significant effects and is not too conservative). In addition, once you know the technique, it is very versatile. Below, I give a few more examples.

6. Getting carried away (again)

One way to check the adequacy of a procedure is to apply it to the classic data sets that have been used in the literature on F2 effects. Most of them come from Raaijmakers et al. (1999).

For instance, Raaijmakers et al. (1999) give the following example (also analyzed by Baayen et al., 2006). It concerns a hypothetical study in which 4 participants take part in a priming study and see 4 items with a short SOA and 4 (different) items with a long SOA.

				TABLE 2					
	Simulat	ed Data for Re	epeated-Measu	rements ANO	VA with Word	ls Sampled Ra	andomly		
		Short	SOA		Long SOA				
Subject	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	
1	546	567	547	566	554	545	594	522	
2	566	556	538	566	512	523	569	524	
3	567	598	568	584	536	539	589	521	
4	556	565	536	550	516	522	560	486	
5	595	609	585	588	578	540	615	546	
6	569	578	560	583	501	535	568	514	
7	527	554	535	527	480	467	540	473	
8	551	575	558	556	588	563	631	558	

For this table Raaijmakers et al. report:

F1(1,7) = 7.41, p = .0297 F2(1,6) = 2.17, p = .1912 minF'(1,10) = 1.68, p = .224

So, how does the multilevel analysis cope? To find out, we again have to write the table in a long form and then run the analysis.

🗰 exan	iple Raai	ijm ake	rs et al	1 - SPSS	Data Editor							
File Edit	View Data	a Transi	form Analyz	e Graphs	Utilities Window Help							
1 : subject		1										
Í T	subject	item	SOA	rt	🔹 SPSS analysis Raaijmakers et al 1 - SPSS Syntax Editor 📃 🔲 🔀	r						
1	1	1	.500	546	File Edit View Data Transform Analyze Graphs Utilities Run Window Help							
2	1	2	.500	567								
3	1	3	.500	547		-						
4	1	4	.500	566	GET FILE = D: Documents and Settings/Marc/Wy Documents/even/Multilevel modeling/example Raaijmakers et al 1.sav.							
5	1	5	500	554	MIXED IT BY Subject Item SOA							
6	1	6	500	545	/FIXED = SOA							
7	1	7	500	594	/METHOD = REML							
8	1	8	500	522	/PRINT = SOLUTION TESTCOV							
9	2	1	.500	566	/RANDOM = INTERCEPT SUBJECT(Item) COVTYPE(UN)							
10	2	2	.500	556	(RANDOM = SOA I SUB ECT(Subject) COVTYPE(IN)							
11	2	3	.500	538	/EMMEANS TABLES (SOA)							
12	2	4	.500	566								
13	2	5	500	512								
14	2	6	500	523								
15	2	7	500	569								
16	2	8	500	524								
17	3	1	.500	567								
18	3	2	.500	598								
19	3	3	.500	568								
20	3	4	.500	584								
21	3	5	500	536								
22	3	6	500	539								
23	3	7	500	589								
24	3	8	500	521								
25	4	1	.500	556	SPSS Processor is ready	1						
26	4	2	.500	565								
27	4	3	.500	536								
28	4	4	.500	550								
29	4	5	500	516								
30	4	6	500	522								
31	4	7	500	560								

These are the results:

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	12.590	2655.786	.000
SOA	1	8.250	1.717	.225

a Dependent Variable: Response Time in Milliseconds.

Estimates of Fixed Effects(b)

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confide	ence Interval
						Lower Bound	Upper Bound
Intercept	563.3125000	12.3815143	9.543	45.496	.000	535.5445293	591.0804707
[SOA=500]	-22.4062500	17.1014526	8.250	-1.310	.225	-61.6356214	16.8231214
[SOA=.500]	0(a)	0	-			-	

a This parameter is set to zero because it is redundant.b Dependent Variable: Response Time in Milliseconds.

These findings [F(1,8.25) = 1.72, p = .225] agree pretty well with those of minF'.

The second example Raaijmakers et al. (1999) gave was a priming study in which the SOA between prime and target was manipulated and in which the items were matched in 4 pairs (called blocks). This is how the data looked like:

	TABLE 4										
	S	imulated Data	for Repeated	Measurements	ANOVA wit	h Matched Iter	ns				
		Short	SOA		Long SOA						
Subject	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4			
1	493	519	513	542	499	525	502	557			
2	562	552	565	591	544	536	533	563			
3	519	558	555	567	575	582	551	587			
4	518	523	514	563	523	565	539	597			
5	567	562	577	595	521	563	559	575			
6	520	534	527	568	512	541	531	559			
7	516	544	513	575	555	569	550	601			
8	525	528	528	559	551	542	529	578			

F1(1,7) = 0.86, p = .385 F2(1,3) = 7.19, p = .075 (by making use of a repeated measures design; see the semantic priming experiment above) minF'(1,3) = 0.77, p = .445

The picture below shows how to do the multilevel analysis:

🗰 exan	nple Raai	ijmake	rs et al .	2 - SPS	SS Data Editor										
File Edit	t View Data	a Transf	orm Analyz	e Graph	hs Utilities Window Help										
2	S 🔍 🖂	al 🔚	? M M		11 II 00										
1 : subject		1													
	subject	item	SOA	rt	var var var var var var var var	T									
1	1	1	.500	493	SDSS analycis Rabiimakors et al 2 - SDSS Syntax Editor										
2	1	2	.500	519	SPSS analysis kaaljinakeis et al 2 - 3r35 Syntax Editor										
3	1	3	.500	513	File Edit View Data Transform Analyze Graphs Utilities Run Window Help										
4	1	4	.500	542											
5	1	1	500	499	GET FILE = 'D'\Documents and Settings\Marc\My Documents\even\Multilevel modeling\evample Raaiimaks										
6	1	2	2500 525												
7	1	3	500	502	MIXED rt BY Subject Item SOA										
8	1	4	500	557	/FIXED = SOA	/FIXED = SOA									
9	2	1	.500	562											
10	2	2	.500	552	/PRANDOM = INTERCEPT I SUBJECT/Item) COVTYPE/UN)										
11	2	3	.500	565	/RANDOM = INTERCEPT SUBJECT(Subject) COVTYPE(UN)										
12	2	4	.500	591	/EMMEANS TABLES (SOA).	_									
13	2	1	500	544		_									
14	2	2	500	536		_									
15	2	3	500	533		_									
16	2	4	500	563											
1/	3	1	.500	519		_									
18	3	2	.500	558		_									
19	5	3	.500	555		-									
20	3	4	.500	567		_									
21	3	1	500	5/5											
22	3	2	500	582		_									
23	3	3	500	551		-									
24	3	4	500	507		-									
25	4	1	.500	518											
20	4	2	.500	523	SPSS Processor is ready										
2/	4	3	.500	514											
20	4	4	.500	503											
29	4		000	525											
31	4	2	000	539		-									
32	4	1	500	597											
33	5	1	.500	567											

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4.721	2637.865	.000
SOA	1	52.000	3.411	.070

a Dependent Variable: Response Time in Milliseconds.

Estimates of Fixed Effects(b)

						95% Confidence Interval		
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound	
Intercept	543.50000	10.814031	5.019	50.259	.000	515.7339981	571.2660019	
[SOA=500]	6.9375000	3.7564648	52.000	1.847	.070	6003980	14.4753980	
[SOA=.500]	0(a)	0						

a This parameter is set to zero because it is redundant.

b Dependent Variable: Response Time in Milliseconds.

Here we see something 'strange': The multilevel analysis is much more 'lenient' than minF' (F(1,52) = 3.41, p = .07). What is happening here? To be honest, I don't know. The only thing I know is that when Baayen et al. (2006) discussed this example, they included an additional random variable, next to participants and items, namely SOA (which is random by participant; the authors do not explain why). If we do so, we get the following:

🔲 exan	nple Raai	ijmake	rs et al 🕯	2 - SPS	iS Data	Editor								
File Edit	View Data	a Transfi	orm Analyz	e Graph	is Utilitie	is Window Help								
2	B	~ <u>"</u> [n 🙀 📲	ř 🔳	₫ ∏ 1	<u>©</u>								
1 : subject		1												
T T	subject	item	SOA	rt	var	var var var var var var var var								
1	1	1	.500	493		SDSS analysis Baaiimakars at al. 2. SDSS Syntax Editor								
2	1	2	.500	519		B SPSS analysis kaaijinakeis et al 2 - SPSS Syntax Euron								
3	1	3	.500	513		File Edit View Data Transform Analyze Graphs Utilities Run Window Help								
4	1	4	.500	542										
5	1	1	500	499		GET FILE = 'D'\Decuments and Settings\Mars\My Decuments\avan\Multilaval medaling\avanna Baaiimaks								
6	1	2	500	525		GET FILE - D. Documents and Settings ward will becoments teven woll liever modeling texample Raaijmake								
7	1	3	500	502		MIXED rt BY Subject Item SOA								
8	1	4	500	557		/FIXED = SOA								
9	2	1	.500	562										
10	2	2	.500	552		/PRINT = SOLUTION TESTCOV /RANDOM = INTERCEPT SUBJECT(Item) COVTYPE(UN) /RANDOM = INTERCEPT SUBJECT(Subject) COVTYPE(UN)								
11	2	3	.500	565										
12	2	4	.500	591		/RANDOM = SOA SUBJECT(Subject) COVTYPE(UN)								
13	2	1	500	544		/EMMEANS TABLES (SOA).								
14	2	2	500	536										
15	2	3	500	533										
16	2	4	500	563										
17	3	1	.500	519										
18	3	2	.500	558										
19	3	3	.500	555										
20	3	4	.500	567										
21	3	1	500	575										
22	3	2	500	582										
23	3	3	500	551										
24	3	4	500	587										
25	4	1	.500	518										
26	4	2	.500	523										
27	4	3	.500	514		5PSS Processor is ready								
28	4	4	.500	563										
29	4	1	500	523										
30	4	2	500	565										
31	4	3	500	539										

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4.857	2600.649	.000
SOA	1	6.714	.862	.385

a Dependent Variable: Response Time in Milliseconds.

Estimates of Fixed Effects(b)

						95% Confidence Interval		
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound	
Intercept	543.50000	11.600214	6.163	46.853	.000	515.2966380	571.7033620	
[SOA=500]	6.9375000	7.4729796	6.714	.928	.385	-10.8872079	24.7622079	
[SOA=.500]	0(a)	0	•	•			-	

a This parameter is set to zero because it is redundant.b Dependent Variable: Response Time in Milliseconds.

Now, the F-value looks much more like what one would expect: F(1,6.7) = .862, p = .385. Apparently in a blocked design you need to define the block as a random variable (I hope to clear this out in a later version).

The final example Raaijmakers et al. (1999) gave was an example in which a Latinsquare design is used. It was a priming study with 3 SOA levels (short, medium, and long) and 12 items that were rotated over the three conditions.

						TA	BLE 7							
				Simulated	l Data fo	r Design	u Using C	ounterba	lanced Li	sts				
			Shor	t SOA		Medium SOA					Long SOA			
Group	Subject	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	
1	1	532	508	522	482	468	496	544	547	475	522	502	484	
	2	542	516	545	483	509	519	588	583	499	535	535	486	
	3	615	584	595	560	542	591	630	617	543	606	560	545	
	4	547	553	584	535	514	555	591	606	538	565	546	527	
		Item 9	Item 10	Item 11	Item 12	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	
2	5	553	598	581	551	619	576	606	561	548	590	614	631	
	б	464	502	485	451	484	479	499	471	447	486	514	523	
	7	481	511	492	472	531	506	542	475	471	510	539	556	
	8	541	588	551	533	582	556	589	515	538	545	601	576	
		Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 1	Item 2	Item 3	Item 4	
3	9	482	530	571	563	501	561	500	506	543	539	558	497	
	10	559	570	632	639	551	592	572	561	617	587	616	549	
	11	462	497	546	538	487	546	491	470	529	508	525	473	
	12	460	463	511	528	457	506	487	453	498	479	512	443	

Raaijmakers et al. calculated a reasonably complicated F-statistic for this design, which yielded F(2,20) = .896, p = .424 (see also below for the 'usual' F1, F2, and minF'). The multilevel analysis gave the following results.

🔲 exan	example Raaijmakers et al 3 - SPSS Data Editor										
File Edit	View Dat	ta Transfori	m Analy:	ze Graphs	Utilities	Window Help					
	s 💷 🖂	a 🔚 🕼	#		5	0					
1 : Group		1									
	Group	subject	item	SOA	rt	var var var var var var var var var					
1	1	1	1	500	532						
2	1	1	2	500	508						
3	1	1	3	500	522	SPSS analysis Raajimakers et al. 3 - SPSS Syntax Editor					
4	1	1	4	500	482						
5	1	1	5	.000	468	File Eult View Data Transform Analyze Graphs Outlides Run Window Help					
6	1	1	6	.000	496						
7	1	1	7	.000	544	GET FILE = 'D:\Documents and Settings\Marc\My Documents\even\Multilevel modeling\example Raaijmake					
8	1	1	8	.000	547						
9	1	1	9	.500	475	MIXED it BY Subject Item SOA					
10	1	1.	10	.500	522						
11	1	1	11	.500	502						
12	1	1	12	.500	484	/RANDOM = INTERCEPT SUBJECT(Item) COVTYPE(UN)					
13	:1	2	1	500	542	/RANDOM = INTERCEPT SUBJECT(Subject) COVTYPE(UN)					
14	1	2	2	500	516	/EMMEANS TABLES (SOA).					
15	1	2	3	500	545						
16	1	2	4	500	483						
17	1	2	5	.000	509						
18	1	2	6	.000	519						
19	1	2	7	.000	588						
20	1	2	8	.000	583						
21	1	2	9	.500	499						
22	1	2	10	.500	535						
23	1	2	11	.500	535						
24	1	2	12	.500	486						
25	1	3	1	500	615						
26	1	3	2	500	584						
2/	1	3	5	500	595						
28	1	3	4	500	560	SPSS Processor is ready					
29	1	3	5	.000	542						
30	1	3	7	000	160						
30	1	3	/	.000	617						
33	1	3	9	.000	543						

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.798	1530.967	.000
SOA	2	119.000	.944	.392

a Dependent Variable: Response Time in Milliseconds.

Estimates of Fixed Effects(b)

						95% Confidence Interval		
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound	
Intercept	533.95833	13.709801	20.084	38.947	.000	505.3678104	562.5488563	
[SOA=500]	4583333	2.0058829	119.000	228	.820	-4.4301819	3.5135152	
[SOA=.000]	2.1250000	2.0058829	119.000	1.059	.292	-1.8468486	6.0968486	
[SOA=.500]	0(a)	0						

a This parameter is set to zero because it is redundant.

b Dependent Variable: Response Time in Milliseconds.

Interestingly, in this analysis Baayen et al. (2006) do not define an extra random variable to capture the repeated measures element. Still, the F-value is quite comparable to the one obtained by Raaijmakers et al., even though it has a much bigger df2 (due to the fact that much less parameters must be estimated).

A concern about the above analysis may be that it doesn't matter that much which analysis you use when the effect is small. So, to see how the different analyses compare when the effects are slightly more interesting, I added 4 ms to the medium SOA condition (half of the data got +4, one quarter +2, and the remaining quarter +6). Given that the variability of the data is quite low, this should suffice to find significance, which is indeed what I found when I ran the usual F1, F2, and minF':

F1(2,18) = 5.481, p = .014 F2(2,18) = 9.426, p = .002minF'(2,34) = 3.456, p = .043 multilevel F(2,119) = 6.742, p = .002

The multilevel analysis gave the following:

🔲 exa	🖩 example Brysbaert 1 - SPSS Data Editor														
File Edi	it View Da	ata Transfori	m Analyz	e Graphs	Utilities 1										
alui.		~ % [6		ret mat		🗈 SPSS analysis Brysbaert 1 - SPSS Syntax Editor 📃 🗌 🔀 🔤									
						File Edit View Data Transform Analyze Graphs Utilities Run Window Help									
6:rt		502													
	Group	subject	item	SOA	rt	var									
1	1	1	1	500	532	GET FILE = 'D:\Documents and Settings\Marc\My Documents\even\Multilevel modeling\example Brysbaert 1.sav.									
2	1	1	2	500	508	MIXED rt BY Subject Item SOA									
3	1	1	3	500	522	/FIXED = SOA									
4	1	1	4	500	482	/METHOD = REML									
5	1	1	5	.000	470	/PRINT = SOLUTION TESTCOV									
6	1	1	6	.000	502	RANDOM = INTERCEPT SUBJECT(IBM) COVTYPE(IN)									
	1	1	/	.000	548	/EMMEANS TABLES (SOA).									
8			8	.000	551										
9	1	1	9	.500	4/5	-									
10	1	1	10	.500	522	-									
12	1	1	12	.000	494	-									
12	1	7	12	.500	5404	-									
14	1	2	2	500	516	-									
15	1	2	3	- 500	545	-									
16	1	2	4	- 500	483	-									
17	1	2	5	000	513										
18	1	2	6	.000	521										
19	1	2	- 7	.000	594										
20	1	2	8	.000	587										
21	1	2	9	.500	499										
22	1	2	10	.500	535	SPSS Processor is ready									
23	1	2	11	.500	535										
24	1	2	12	.500	486										
25	1	3	1	500	615										
26	1	3	2	500	584										
27	1	3	3	500	595										
28	1	3	4	500	560										
29	1	3	5	.000	546										
	1	3	6	.000	595										
31	1	3	7	.000	632										
32	1	3	8	.000	623										
33	1	3	9	.500	543										

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.797	1538.707	.000
SOA	2	119.000	6.742	.002

a Dependent Variable: Response Time in Milliseconds.

Estimates of Fixed Effects(b)

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	533.95833	13.709200	20.082	38.949	.000	505.3689197	562.5477469
[SOA=500]	4583333	2.0020390	119.000	229	.819	-4.4225705	3.5059039
[SOA=.000]	6.1250000	2.0020390	119.000	3.059	.003	2.1607628	10.0892372
[SOA=.500]	0(a)	0					

a This parameter is set to zero because it is redundant.

b Dependent Variable: Response Time in Milliseconds.

Something else I tried, was see what happens if one participant gets an extra 100 ms on all items (see the example above for the slow participant). If the underlying reasoning of the technique is what it claims to be, then this should have no effect on the F-statistic for SOA, because the change can easily be captured by a different intercept for the participant involved. So, we should get rid of the requirement to introduce between-items Latin-square variables or the necessity to work with z-scores. This is exactly what happened, as can be seen in the following tables:

🗰 exan	nple Bry	sbaert 2	- SPSS	Data Ed	itor						
File Edit	: View Da	ita Transfori	m Analyz	te Graphs	Utilities	Window	w Help				
- ALITIA			aa I 🚛			al					
	2 1		94			0					
1 : Group		1					SPSS analysis brysbaert 2 - SPSS Syntax Editor				
	Group	subject	item	SOA	rt	var	File Edit View Data Transform Analyze Graphs Utilities Run Window Help				
1	1	1	1	500	632						
2	1	1	2	500	608		GET FILE = 'D'\Documents and Settings\Marc\My Documents\even\Multilevel modeling\example Bryshaert 2 say				
3	1	1	3	500	622						
4	1	1	4	500	582		MIXED it BY Subject Item SOA				
5	1	1	5	.000	569		/FIXED = SOA				
5		1	5	.000	602						
	1	1	/	.000	648		/RANDOM = INTERCEPT SUBJECT(Item) COVTYPE(UN)				
- 0	1	1	0	.000	601 E7E		/RANDOM = INTERCEPT SUBJECT(Subject) COVTYPE(UN)				
10	1	1	10	.000	670		/EMMEANS TABLES (SOA)				
11	1	1	11	500	602						
12	1	1	12	500	584		-				
13	1	2	1	- 500	542		-				
14	1	2	2	500	516						
15	1	2	3	500	545						
16	1	2	4	500	483						
17	1	2	5	.000	513						
18	1	2	6	.000	521						
19	1	2	7	.000	594						
20	1	2	8	.000	587						
21	1	2	9	.500	499						
22	1	2	10	.500	535						
23	1	2	11	.500	535						
24	1	2	12	.500	486		SPSS Processor is ready				
25	1	3	1	500	615						
26	1	3	2	500	584						
27	1	3	3	500	595						
28	1	3	4	500	560						
29	1	3	5	.000	546						
30	1	3	5	.000	595						
- 31	1	3	/	.000	632						
32	1	3	8	.000	6/3						
a second s					- 144 T						

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.743	1383.900	.000
SOA	2	119.000	6.742	.002

a Dependent Variable: Response Time in Milliseconds.

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	542.29167	14.673791	18.978	36.956	.000	511.5766442	573.0066891
[SOA=500]	4583333	2.0020390	119.000	229	.819	-4.4225705	3.5059039
[SOA=.000]	6.1250000	2.0020390	119.000	3.059	.003	2.1607628	10.0892372
[SOA=.500]	0(a)	0				-	

Estimates of Fixed Effects(b)

a This parameter is set to zero because it is redundant.

b Dependent Variable: Response Time in Milliseconds.

Finally, I wanted to see what happens when 1 observation in Raaijmakers et al.'s table got a much higher value (participant 1, item 5 +120 ms). Will this turn the multilevel F-statistic into a spurious significance?

	nalo Pro	change 2	CDCC	Data Ed	liter		
i exa	прте вгу	spaert 5	- 3535	Dala Cu	ntor		
File Edi	t View Da	ita Transforr	m Analya	te Graphs	Utilities	Window He	ap
B	5 🔍 🗠	a 🔚 📭	M 📲			0	
5 : rt		588					
1	Group	subject	item	SOA	rt	var	SCDSS analysis Brythaart 2 SDSS Syntax Editor
1	1	1	1	500	532		a sess analysis biysbaert s - sess syntax Eulton
2	1	1	2	500	508		File Edit View Data Transform Analyze Graphs Utilities Run Window Help
3	1	1	3	500	522		
4	1	1	4	500	482		CET FILE = "D'Decumente and Settings/Marc/My Decumente/avan/Multilaval modeling/avangle Bruchaet 3 cov
5	1	1	5	.000	588		
6	1	1	6	.000	496		MIXED rt BY Subject Item SOA
7	1	1	7	.000	544		/FIXED = SOA
8	1	1	8	.000	547		METHOD = REML
9	1	1	9	.500	475		PAINT = SOLUTION TESTCOV PAINTOM = INTERCEPT (SUBJECT/Item) COVTYPE(IN)
10	1	1	10	.500	522		RANDOM = INTERCEPT SUBJECT(IEII) COVTYPE(IN)
11	1	1	11	.500	502		/EMMEANS TABLES (SOA).
12	1	1	12	.500	484		
13	1	2	1	500	542		
14	1	2	2	500	516		
15	1	2	3	500	545		
16	1	2	4	500	483		
17	1	2	5	.000	509		
18	1	2	6	.000	519	1	
19	1	2	7	.000	588		
20	1	2	8	.000	583		
21	1	2	9	.500	499		
22	1	2	10	.500	535		
23	1	2	11	.500	535		
24	1	2	12	.500	486		
25	1	3	1	500	615		
26	1	3	2	500	584		SPSS Processor is ready
27	1	3	3	500	595		
28	1	3	4	500	560		
29	1	3	5	.000	542		
30	1	3	6	.000	591		
31	1	3	7	.000	630		
32	1	3	8	.000	617		

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.421	1610.356	.000
SOA	2	119.000	2.055	.133

a Dependent Variable: Response Time in Milliseconds.

						95% Confidence Interval		
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound	
Intercept	533.95833	13.436372	19.984	39.740	.000	505.9291461	561.9875206	
[SOA=500]	4583333	2.7740527	119.000	165	.869	-5.9512348	5.0345681	
[SOA=.000]	4.6250000	2.7740527	119.000	1.667	.098	8679014	10.1179014	
[SOA=.500]	0(a)	0				-	-	

a This parameter is set to zero because it is redundant.

b Dependent Variable: Response Time in Milliseconds.

The obtained F-value [F(2,119)=2.055, p < .133] compares favorably to what happens with F1 and F2 (although not to minF', which is a good reminder that the F1 x F2 criterion may give the wrong impression):

F1(2,18) = 2.739, p = .092 F2(2,18) = 3.061, p = .072 minF'(2,36) = 1.44, p = .249

Finally, the multilevel design is not limited to a single IV. Locker et al. (2007) give an example of an LDT experiment in which the effects of phonological neighborhood frequency and semantic neighborhood size were measured. This is their code (which can easily be adapted).

```
MIXED rt BY Subject Item Freq Size

/FIXED = Freq Size Freq*Size

/METHOD = REML

/PRINT = SOLUTION TESTCOV

/RANDOM = INTERCEPT | SUBJECT(Item) COVTYPE(UN)

/RANDOM = INTERCEPT | SUBJECT(Subject) COVTYPE(UN)

/EMMEANS TABLES (Freq*Size).
```

In summary, I am becoming more and more convinced that multilevel modeling is the way forward. The analyses are easier than than the F1, F2, and minF' calculations and they seem to be of a higher quality. In the final section, I refer to one more advantage of the multilevel approach.

7. Beyond dichotomizing

For someone with a bit of experience in analyzing psycholinguistic data, the idea of simultaneously controlling for item and participant variation must ring a bell. In 1990, Lorch and Myers published an article on how to do a proper linear regression in a repeated measures design. The problem is analogue to the one discussed in Figure 1, although now it involves generalization over participants.

The problem is illustrated in Table 4, where the results are shown for 6 participants on 10 items that vary in log10(frequency).

	LogFreq	RTpart1	RTpart2	RTpart3	RTpart4	RTpart5	RTpart6
1	.25	900	625	601	706	821	489
2	.50	850	654	609	652	812	512
3	.75	800	699	614	717	845	497
4	1.00	750	599	610	642	854	468
5	1.25	700	652	630	713	823	501
6	1.50	650	603	624	695	832	466
7	1.75	600	631	637	689	861	484
8	2.00	550	622	629	664	815	503
9	2.25	500	669	643	703	769	498
10	2.50	450	599	641	678	804	527

Table 4 : Example of regression data in a design with a repeated measure (LDT to10 words varying in frequency).

If we average the data over the 6 participants and calculate the regression analysis, we get:

RT = 702 - 33.5 LogFreq (LogFreq: t(8) = -7.588, p < .001, R² = .88).

A look at Table 4 makes clear where this huge frequency effect comes from (and how things can go pear-shaped). Only one of the participants (i.e., part1) shows a substantial linear frequency effect. All the others show either no effect or even a slight opposite effect. Unfortunately, this variability is lost when the regression is based on the mean RT over participants.

To counter this problem, Lorch & Myers (1990) suggested to do a separate analysis per participant and then to run a t-test on the regression weights obtained. So, they would do the following calculations:

Part1 : 950 – 200 LogFreq Part2 : 651 – 11.3 LogFreq Part3 : 599 + 18.1 LogFreq Part4 : 687 – .9 LogFreq Part5 : 843 – 13.9 LogFreq Part6 : 485 + 7.0 LogFreq

Table 3

A simple one-sample t-test reveals that in the Lorch & Myers (1990) analysis, the effect of LogFreq is not significant (t(5) = -.996, p = .365).

Ever since many psycholinguists have happily spent days calculating regression weights of individual participants and running one-sample t-tests on them, even though apparently there is a simpler way to get at it directly from the ANOVA table.

If you want to have a go at this type of analysis, here is the example Lorch & Myers worked with in their article. It deals with sentence reading times as a function of the rank order of the sentence, the number of words in the sentence, and the number of new words in the sentence.

OBSERVATIONS

153

Subjects' Reading Times and Values of Predictor Variables for Each Sentence of the Experimental Text

					•		-								_
_	SNT	SP	WRDS	NEW	SBJ 1	SBJ 2	SBJ 3	SBJ 4	SBJ 5	SBJ 6	SBJ 7	SBJ 8	SBJ 9	SBJ 0	
	1	1	13	1	3429	2795	4161	3071	3625	3161	3232	7161	1536	4063	Ĩ
	2	2	16	3	6482	5411	4491	5063	9295	5643	8357	4313	2946	6652	
	3	3	9	2	1714	2339	3018	2464	6045	2455	4920	3366	1375	2179	
	4	4	9	2	3679	3714	2866	2732	4205	6241	3723	6330	1152	3661	
	5	5	10	3	4000	2902	2991	2670	3884	3223	3143	6143	2759	3330	
	6	6	18	4	6973	8018	6625	7571	8795	13188	11170	6071	7964	7866	
	7	7	6	1	2634	1750	2268	2884	3491	3688	2054	1696	1455	3705	

Note: SNT = sentence; SP = serial position of sentence; WRDS = number of words in sentence; NEW = number of new arguments in sentence; SBJ = subject.

This is the analysis Lorch & Myers reported:

Subject	SP	WORDS	NEW
1	0.23124	0.39103	0.22161
2	0.30533	0.43415	0.34637
3	0.20637	0.40360	25294
4	0.48300	0.50203	27683
5	-0.06210	0.28778	0.92680
6	1.10982	0.80850	23336
7	0.25448	0.57498	0.79643
8	-0.33147	0.11341	0.33124
9	0.66786	0.50078	0.16320
10	0.46921	0.56964	50621
М	0.33337	0.45859	0.15163
SE	0.12417	0.05855	0.14982
t	2.6849	7.83289	1.01210

Table 4 Regression Coefficients From Individual Analyses of

Note. sp = serial position of sentence; words = number of words in sentence; NEW = number of new arguments in sentence.

From this they concluded that the serial position of the sentence and the number of words were significant predictors of reading time, but not the number of new words.

Van den Noortgate and Onghena (2006) used this example to show how much easier multilevel programming is. The nice thing about the MIXED function is that it not only works with discrete variables but also with continuous variables (the only thing you have to change is to use WITH instead of BY in the model specification). This is the program Van den Noortgate & Onghena used:

No.						
🗰 exa	mple Lor	rch & Mye	ers - SP	SS Data	Editor	
File Ed	lit View Da	ita Transfori	n Analyz	e Graphs	Utilities Wi	indow Help
			44 ×	et ma		1
senten						
	sentence	subject	words	new	RT	SPSS analysis Lorch & Avers - SPSS Syntax Editor
1	1	1	13	1	3.429	
2	1	2	13	1	2.795	File Edit View Data Transform Analyze Graphs Utilities Run Window Help
3] 1	3	13	1	4.161	
4] 1	4	13	1	3.071	
5	1	-5	13	1	3.625	GET FILE = D-\Documents and Settings\Marc\My Documents\even\Multilevel modeling\example orch & Myers sav
6] 1	6	13	1	3.161	
7] 1	7	13	1	3.232	MIXED rt WITH Sentence words New
8] 1	8	13	1	7.161	/FIXED = Sentence Words New
9	1	9	13	1	1.536	PRINT = SOLUTION TESTCOV
10	1	10	13	1	4.063	/RANDOM = INTERCEPT Sentence Words New [SUBJECT(Subject) COVTYPE(ON).
11	2	1	16	3	6.482	
12	2	2	16	3	5.411	
13	2	3	16	3	4.491	
14	2	4	16	3	5.063	
15	2	5	16	3	9.295	
16	2	6	16	3	5.643	
17	2	7	16	3	8.357	
18	2	8	16	3	4.313	
19	2	9	16	3	2.946	
20	2	10	16	3	6.652	1 –
21	3	1	9	2	1.714	1
22	3	2	9	2	2.339	SPSS Processor is ready
23	1 3	3	9	2	3.018	

with the following results:

Estimates of Fixed Effects(a)

						95% Confide	ence Interval
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	-2.586950	.7425953	19.755	-3.484	.002	-4.1372114	-1.0366896
sentence	.3333728	.0989789	36.617	3.368	.002	.1327516	.5339941
words	.4585893	.0680731	36.617	6.737	.000	.3206113	.5965673
new	.1516299	.2560739	36.617	.592	.557	3674087	.6706684

a Dependent Variable: Reading time.

When we do the same analysis on on our simple example with the word frequency data, we get

🗖 line	ar regre	ssion wo	ord freq l	ong versio	on - SPSS D)ata Editor										- 6 X
File Ed	it View D	ata Transfo	rm Analyze	Graphs Utili	ties Window	Help										
	a 🗉 🗠				<u>s</u>											
1 : Particin	iant	1														
	Deuticinent	LogErog	PoodTim	vor 1	100-00-00-00-00-			9 90 90 90 M 9		0.0000000000000000000000000000000000000		STORE STORE			Vor	Unr 🔺
1	Participant	Lugrieg 25	900	var	SPSS ana	lysis linea	regress	ion log fi	equenc	y - SPSS S	iyntax I	Editor			Var	var -
2	1	.50	850		File Edit Viev	v Data Trans	form Analyz	e Graphs	Jtilities Ru	n Window I	Help					
3	1	.75	800				6 46 1	ON EN !	1							
4	1	1.00	750													
5	1	1.25	700		GET FILE = 'D'	FILE = 10 \Documents and Sattings\Marc\My Documents\avan\Multilaval modeling\linear regression word free long version say										
6	1	1.50	650		oernee o.	Decourrente and	Contingo inte	io ing boodi	011101010101111		ing into a ri	-greeelen ne	ia noq iong i	ioroion.our.		
7	1	1.75	600		MIXED ReadTir	n WITH LogFreq										
8	1	2.00	550		/FIXED	LegFreq	ERTOON									
9	1	2.25	500		/RAND	OM = INTERCE	PT LogFreg L	SUBJECT(Pa	rticipanth CC							
10	1	2.50	450													
12	2	.25	625 GE 4	<u> </u>											-	<u></u>
12	2	.50	#C0													
14	2	1.00	599												-	
15	2	1.25	652													
16	2	1.50	603													
17	2	1.75	631													
18	2	2.00	622													
19	2	2.25	669													
20	2	2.50	599													
21	3	.25	601													
22	3	.50	609						SPSS Proce	ssor is ready	5				11	
23	3	.75	614					-							_	
24	3	1.00	610					1								
20	3	1.25	634												_	
20	3	1.50	637				-						-		-	
28	3	2.00	629					13	12			2		0		2
29	3	2.25	643					0		· ·				2		
30	3	2.50	641		1											
31	4	.25	706	2	9	8	2	9	0	0		2	2	0		2
32	4	.50	652											1		
33	4	.75	717													
34	4	1.00	642		2											
35	4	1.25	713					3	5			[<u></u>	2		
36	4	1.50	695													
37	4	1.75	689			1				1		<u> </u>				
38	4	2.00	664					0	0	-						
39	4	2.25	/03				_									
40	4 	2.5U ⊃£	871				-	0	17	1				-	-	
41	5	.25	812	-			-									+
42	5	.50	845		1		-		-	1		ł		1	1	
40	5	1.00	, 954													<u>+</u>
< > \c	Data View /	Variable V	iew /						once n							
			Internet	1	1000	-		Transfer	SPSS Proce	essor is ready		T	-			
💾 sta	art	🔾 3 In 🔻	💾 The I	🔄 Multil	Micro	4 S	👻 👹 Analy.	🗐 2 0	🎽 🐻	Van D	Calcul	🛃 Adob	FR 🤇	Norton (S 🛛 🖉 S	17:04

Estimates of Fixed Effects(a)

						95% Confide	ence Interval
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	702.42222	68.77472	5.000	10.213	.000	525.6311788	879.2132657
LogFreq	-33.50708	33.646936	5.000	996	.365	-119.9992728	52.9851314

a Dependent Variable: ReadTim.

8. Conclusion

There is an ongoing complaint among teachers and lecturers that students nowadays know less than students some time ago (despite the Flynn-effect). Until recently I thought this was because teachers and lecturers were good students themselves and therefore have a biased view of the motivation and the level of knowledge of their cohort (as they did not tend to interact a lot with the 'bad' students). A few months ago, however, I came across an article in which an educational psychologist gave another explanation. According to him, teachers in particular see the lack of knowledge in students for what *they themselves* know well on the basis of their education (e.g., history, geography, correct spelling, algebra, elementary statistics, ...), but they fail to notice the knowledge pupils/students have that is not shared by teachers/lecturers. When it comes to acquiring new knowledge and skills, teachers are no better than students if the immediate use of the knowledge is not obvious.

This view has crossed my mind a few times in the past couple of days: Is it possible that we keep on clutching to the familiar F1 and F2, because we've learned to calculate them in our undergraduate studies (in my case even by hand)? My present journey most certainly has convinced me that I seem to have missed a few steps in current statistical sophistication. It certainly is an incentive to explore the *lme4* package (http://cran.r-project.org), which has many more goodies and possibilities than what is on offer in SPSS (Baayen, 2007; Baayen et al., 2006). The present review shows that a better understanding of multilevel analysis techniques (or mixed-effects techniques) is likely to be rewarding, although it is amazing how much is already available in the statistical program we use daily, at no larger clicking cost than we are doing now (often quite the contrary as I have found out)!

9. References

Baayen, R.H. (2007). Analyzing linguistic data: A practical introduction to statistics. Cambridge: Cambridge University Press (in press).

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2006). *Mixed-effects modeling with crossed random effects for subjects and items*. Available on the internet (copy the title in google).

Clark, H.H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.

Locker, L., Hoffman, L., & Bovaird, J. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*.

Lorch, R.F., & Myers, J.L. (1990). Regression analysis of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 149-157.

Pollatsek, A., & Well, A.D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 785-794.

Raaijmakers, J.G.W. (2003). A further look at the "language-as-fixed-effect fallacy'. *Canadian Journal of Experimental Psychology*, *57*, 141-151. Raaijmakers, J.G.W., Schrijnemakers, J.M.C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416-426.

Van den Noortgate, W., & Onghena, P. (2006). Analysing repeated measures in cognitive research: A comment on regression cofficient analyses. *European Journal of Cognitive Psychology*, *18*, 937-952.