**An Extension of the QWERTY Effect: Not Just the Right Hand, Expertise and Typability Predict Valence Ratings of Words**

Reviewed by Daniel Casasanto

The authors report 2 experiments designed to replicate and extend the QWERTY Effect: The finding that words typed with more letters from the right side of the keyboard tend to be more positive in valence (Jasmin & Casasanto, 2012; hence J&C). They tested for relationships between the emotional valence of words and variables that could affect typing fluency. Overall, the clearest result was that the QWERTY effect was shown across several corpora, leading the authors to conclude that, “the original QWERTY effect was replicable across a large body of various types of stimuli (verbs, Twitter, category norms), with much the same size of effect as Jasmin and Casasanto (2012) published.”

I would \*love\* to see this claim published in JML: a strong replication of a controversial study. But, unfortunately, I see nothing in this paper that’s publishable. The paper is problematic for several reasons. The most irremediable problems are (1) the data from Experiment 1, (2) the methods and data from Experiment 2, and (3) I am sad to report, what I believe to be some matters of questionable academic integrity concerning both experiments.

I reviewed an earlier version of this paper, submitted to another journal, reporting an earlier analysis of Expt. 2. Although my first review was critical of the experiment and other aspects of the paper that were (and still are) problematic, it was intended to be sympathetic and encouraging to the authors. This review is somewhat less sympathetic, for reasons that should become obvious.

**Concerns about Scholarship and Ethics**

**1. The authors reported analyses that J&C had already done, in some of the same corpora. They omitted any mention of this fact, even though they had been informed of their oversight previously.**

In the first experiment, the authors sought to extend the QWERTY effect by testing for the effect of a variable related to typing fluency: Hand alternations, or “switches.” They motivated the experiment with two main predictions (and with a third prediction that combines 1 and 2). From pg. 7:

1. “RSA [i.e., J&C’s “Right Side Advantage” metric] should be a significant predictor of valence ratings, even after controlling for word length and average letter frequency[.]”
2. “[We] expect to find that increased hand switches are positively related to ratings of valence because words that are typed on alternating hands are easier to type.”

There’s a problem with motivating the study with these two predictions: J&C (2012) already tested both of them explicitly, in 5 corpora. In all 5 of J&C’s corpora, RSA predicted valence when word length and letter frequency were controlled: This was, in fact, the main finding of J&C’s paper. Also, in all 5 corpora, Hand Alternations (i.e., “switches”) showed no significant relationship with valence.

So, were the present authors taking issue with J&C’s analyses? Or were they approaching the same questions in a substantively different way? No. They never mentioned that J&C had already done these analyses and reported their results.

Could the authors’ failure to discuss the fact that J&C had already conducted these analyses be an innocent oversight? This is very unlikely with respect to the frequency and length analyses, and impossible w. r. t. the Hand Alternations analyses.

**Frequency and length control analyses:**

The authors’ first two analyses used the corpora ANEW and AFINN. These were also the first two corpora analyzed in J&C’s paper, which reported the following:

J&C ANEW:

“A further analysis was conducted to control for possible effects of word length and for the frequency with which individual letters are used in each language (letter frequency). RSA remained a significant predictor of valence when word length, letter frequency, language, and their interactions were controlled (b=0.057, Wald X2(1)=6.95, p=0.008). [Note: Language was also included in our model because J&C tested ANEW in 3 languages.]

J&C AFINN:

“Finally, as in Experiment 1, RSA remained a significant predictor of valence when word length, letter frequency, and their interaction were controlled (b=0.06, Wald X2(1)=19.06, p=0.001].”

In J&C’s brief report, it would be hard for any earnest reader to miss these analyses – certainly any reader earnest enough to conduct follow-ups in 6 corpora should have noticed them.

**Hand alternation analyses:**

There is no question that the authors were aware that J&C had tested for effects of hand alternations (switches) and found no effects. If they somehow missed our discussion of these analyses in our paper, they could not have missed the following comment from my last review:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Jasmin & Casasanto already conducted analyses of hand alternation and finger repetition.**

The fact that the present authors decided to do an experiment to find out whether the rate of Hand Alternations correlates with word valence suggests they may have overlooked the following paragraphs in J&C’s General Discussion:

“[Our] proposal is broadly consistent with previous research showing influences of typing fluency on preference judgments for meaningless letter strings (e.g., Beilock & Holt, 2007; Van den Bergh et al., 1990). However, previous studies have focused on different sources of typing fluency, such as finger repetition. For example, skilled typists prefer pairs of letters typed with different fingers (“f–j”) over pairs typed with the same finger during standard touch typing (“f–v”; Beilock & Holt, 2007). **In exploratory analyses, we found no significant relationship between the number of finger repetitions in a word and its valence,** **nor was there any relationship between valence and the number of hand alternations used when typing a word—for any of the corpora we analyzed.**

**These other sources of typing fluency are orthogonal to the number of right-side and left-side letters in a word, and the effects we report here remain significant when both finger repetition and hand alternation are controlled.”** (J&C, pg.503)

So, J&C already tested for the effect of hand alternation on valence that the authors report here – 5 times – and found that it was not significant in any of the corpora. Furthermore, J&C found that their predicted effect of RSA on valence remained significant when these other sources of fluency were controlled.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

In the original submission, I was willing to accept that the authors’ failure to acknowledge our previous tests of the Hand Alternation hypothesis was a simple oversight. In the revised paper, however, this is not credible. The fact is: The authors did almost exactly the same analyses that J&C reported, on some of the same corpora – knowingly – but presented these analyses as original.

Furthermore, they claimed that their impetus for conducting the hand alternation analyses was their reading of Beilock & Holt (2007). This also seems implausible, given that Beilock & Holt never discussed hand alternation – they conducted a finer-grained “finger repetition” task. By contrast, J&C conducted both hand alternation and finger repetition analyses, and cited Beilock & Holt as part of the motivation for the latter (see above).

This leads me to a second point of concern about ethical conduct.

**2. The authors are not reporting the effect of Finger Repetition, a highly relevant variable, which should have been a key predictor of word valence, and which they included in their statistical model in the original submission.**

The particular typing fluency studies that partly motivated J&C’s QWERTY Effect study used Finger Repetition as the predictor of word valence – not hand alternation. (Although these variables are not independent, alternating hands requires switching fingers, but switching fingers does not require alternating hands.)

In their original submission, the authors reported that they included Finger Repetition as a “control variable” in their analyses (no justification given). In my earlier review, I pointed out that this choice would be very hard to justify: Studies like Beilock & Holt’s show that finger repetition predicts typists’ valence ratings for letter pairs, as the authors are aware. As such, their stated hypothesis clearly implies that any variable that links typing fluency with valence should be a predictor variable, not a control variable.

Put another way, I see no way for the authors to justify analyzing hand alternation as a predictor of word valence but not analyzing finger repetition as a predictor of word valence: their hypothesis implies that either should work, in principle. And if they were going to choose only one of these variables to test *a priori*, the literature that the authors cite indicates that it should be Finger Repetition.

I made this suggestion to the authors in my last review: Since finger repetition was already in their model, they should report its effects, and evaluate their hypothesis in light of them. They didn’t do this. Instead, they removed Finger Repetition from their model entirely. Why?

It’s natural to wonder whether the authors were aware that Finger Repetition should have been used as a predictor– after all, their analysis was inspired by Beilock & Holt’s finger repetition study. Could it be that Finger Repetition was classified as a “control variable” in the original submission, and omitted entirely from this submission, because this predictor did not show the predicted effect on valence?

Analyzing Finger Repetitions in ANEW and AFINN would have been just as unmotivated as analyzing Hand Alternations was, since J&C already reported there was no effect of either of these variables in these corpora (or in 3 others). My point is that it’s very hard to justify analyzing one of these variables but not the other, or to justify classifying Finger Repetitions as a control variable, and then omitting it post-hoc.

Similarly, “Typing Speed” was treated as a control variable in the original submission, but promoted to a predictor variable here. Why?

In any case, deciding post-hoc whether a variable is a predictor or a control -- and failing to report a (highly relevant) variable that has been coded and analyzed -- these are ethical no-no’s that promote Type I error: what Simmons, et al., (2011, *Psych Science*) referred to as abuses of “researcher degrees of freedom.”

**Concerns about the results**

**1. The effects of greatest interest show no reliable effects.**

**Analysis of Hand switching**

The effect of greatest interest to the authors, as stated in the Introduction, was the main effect of Hand Alternations (i.e., Switch) on word valence. It is not clear why the authors chose to do a study to test this effect, given that J&C already reported testing it in 5 corpora and finding no significant results.

Over the 2 experiments, the authors conducted 9 tests of the main effect of Switch: in 5 existing corpora, in a combination of these corpora, and in a 6th constructed corpus (Ex 2) that they analyzed 3 different ways.

The predicted effect of hand alternations (switches) was significant in only 3 of these 9 analyses. In other words, the analyses failed to confirm the authors’ main hypothesis twice as often as they confirmed it. The slope of the effect was positive for 4 of the analyses (consistent with the author’s prediction), but negative for 5 of the analyses. That is, the slopes flipped from one corpus to the next, but most of the analyses showed a main effect of Switch that went in the wrong direction.

There is no justification for interpreting these data as support for an effect of Hand Alternation on word valence. This is a non-effect, consistent with J&C’s analyses.

**Analysis of RSA x Switching**

The authors also tested for an interaction of Switching with RSA. This interaction effect was significant in only 3 of the 9 analyses (table 1-3). Like the main effect of switching, this interaction effect was non-significant twice as often as it was significant.

And yet, the authors discuss this interaction as illuminating the QWERTY effect. How do they arrive at this conclusion? By subdividing the data post-hoc, the authors observed an intriguing-seeming pattern, pieces of which were significant in two of their analyses. From pg. 12:

“Number of hand switches was not a significant predictor of valence. However, the interaction between switches and RSA was significant [in the analysis of all corpora combined]. To examine this effect, we coded words as more right-handed (+1SD RSA), equally right-left (RSA = 0) or more left-handed (-1SD RSA) in a simple slopes analysis. Left-handed words showed a non-significant positive effect of switches on valence (b = .014), **while right-handed words showed a significant negative effect of switches on valence (b = -.048).** This finding indicated that if words are being typed on the right hand, generally, we like words that do not switch back and forth, while words typed on the left hand are more pleasant if they switch back to the right hand.”

Is this evidence for an interesting effect of hand alternation, which adds complexity to the QWERTY effect, providing impetus to look beyond the effect of RSA? No, sadly, for three reasons.

First, this effect is composed of two opposing patterns: A non-significant positive slope in the left hand (consistent with the proposal that switching increases valence), and a significant negative slope in the right hand, which suggests, prima facie, that switching hands leads to a *decrease* in words’ valences (see bold passage, above). So, the only significant effect here goes exactly *backwards* of the authors’ prediction that increased switching should correlate with positive valence.

Should we interpret this pattern described on pg. 12 as evidence for a negative relationship between switching and valence, and therefore a disconfirmation of their hypothesis? No, for several reasons. The simplest reason is that they report a similar analysis for Expt 2 (pg. 17) in which the significance of the positive and negative slopes is reversed.

A second reason why this paradoxical-seeming pattern should not be interpreted as a complex, unpredicted, self-contradictory effect of switching on word valence: The knot unravels if we consider this pattern may not be a confusing effect of hand switching, but rather a straightforward consequence of people’s preference for letters on the right side of the keyboard.

The authors offered an astute description of this post-hoc finding: For words typed primarily on the left of the keyboard, people want to switch – that is, they want to switch to using letters on the right side. But for words typed primarily on the right of the keyboard, people do not want to switch – they want to keep using letters on the right side. In other words, there is a Right Side Advantage!

I suspect all that this interaction of Switching with RSA shows is that people prefer to type on the right side of the keyboard: thus, words that start on the left and switch to the right (making the words less left-biased) tend to get more positive valence ratings, as do words that start on the right and stay there (maintaining their right bias). In summary, this apparent influence of switching is, most likely, just a consequence of people’s preference for words typed with more letters on the right than on the left. The RSA is the only principle needed to predict this pattern.

A third, much simpler reason why the RSA x Switching effect is not interpretable in the way the authors hoped (or at all): It is not reliable. It was not significant in *any* of the 5 existing corpora that they tested. It was only significant in 2 data sets: the combination of the 5 corpora, and the new corpus created for Expt 2. And as I’ll explain below, neither of these data sets should be analyzed, at all.

Combination of the 5 existing corpora.The motivation to combine these corpora is clear: the predicted effect of Switch was only significant in 1 of the 5 the individual corpora. Unfortunately, there is no way to combine these corpora sensibly. The data were collected under wildly different conditions, with different populations, and very different rating scales: How can we know how ratings on one scale correspond to ratings on the others?

Even if the scales were normalized, this would not solve the problem of incomparability. Although it seems obvious, the corpora contained different collections of words. Well, this matters for how Ss use the scale you give them! Ss calibrate their ratings to the range of words they encounter, and the ranges differed drastically across corpora (e.g., AFINN includes lots of profanity; ANEW includes no profanity; Dodds included lots of bizarre unpronounceable strings like “#ff”). If a corpus includes many words with extreme valences, the ratings of words with moderate valences may get compressed in the middle of the scale; if there are few extreme words, Ss make finer distinctions between moderate words, ratings of which may expand to fill the range of values. I can see no sensible way to combine these corpora.

I’ll note that J&C combined 3 of their corpora, but that was different: It was the same corpus translated and re-rated in 3 languages. Translation equivalents of the same set of words were rated using the same scale. And the goal of combining them was *not* to try to squeeze some power out of failed analyses – it was to determine whether the magnitude of the QWERTY effect differed across languages.

The corpus created for Expt 2**.** I gave an extensive critique of Experiment 2 in my last review, which I will not repeat here. Suffice it to say that, even though the authors obtained some significant p-values in the analyses reported here, none of the results of Expt 2 can be interpreted meaningfully. Quickly, 3 reasons why not:

1. We have totally insufficient information about how the corpus was constructed – especially in the present version of the paper, which omits many key details that were included in the last version: the corpus was originally constructed to have 24 bins of 10 stimuli each, half words half pseudowords, varying in RSA. How bias and expectancy effects were avoided during sample construction was not discussed in the paper.
2. Even if sampling procedures were unbiased, the resulting corpus is not appropriate for the present purposes, for two reasons. First, the hand-tailored balanced distribution of RSA’s (i.e., 10 words in each RSA bin) bears no resemblance to the distribution of RSA’s in any preexisting corpus, or to the distribution of RSA’s in words as people ordinarily use them. Second, the N’s are tiny. 120 real words (an order of magnitude smaller than J&C’s smallest corpus) and 120 pseudowords (2 orders of magnitude smaller than J&C’s pseudoword corpus). For these reasons, it would be very hard to make an argument for the generalizability of any results obtained in this corpus.
3. The results changed COMPLETELY from the previous submission to this one. The original paper claimed to be a failure to replicate the QWERTY effect -- in the same data that are now called Expt 2, and which now show highly significant effects of RSA in all analyses. The present analyses are far more appropriate than the initial analyses, so I have more confidence in these analyses than in the last set submitted. But it is not confidence inspiring that the results totally reversed from one draft to the next.

**Analysis of Typing speed**

In Expt 2, the authors collected Typing Speed as an index of typing expertise. When they claim to have found effects of “expertise” on valence in the title, abstract, etc. they are referring to a 3-way interaction of Typing Speed x Switch x RSA. This interaction was not predicted – nor has it been adequately explained. (Neither was the even more complex 4-way interaction with speed.)

The authors explore this effect by binning RSA and Typing Speed, and finding that, within negative RSA real words only (an unlicensed post-hoc sub-division of the data), the effect of Switch on Valence was modulated by Typing Speed. However, based on descriptions of the stimulus construction in the methods sections of this submission and the last, I infer that there were only 40 real words with negative RSA in the entire sample – far too few to support a generalizable claim that either Switches or Typing Speed is related to Valence, especially given the null effects of Switches in the other larger corpora.

But the problems with interpreting this effect as support for the authors’ conjecture about expertise get much worse. Even if the interaction of Typing Speed and Switches could be trusted -- it goes the wrong way. Poor typists showed the *strongest* effect of Switches, followed by mediocre typists and then expert typists. The authors cite studies showing that effects of bodily experience are often stronger (or are only present) in experts who have more motor experience in the domain of study. Specifically, Beilock & Holt (2007) found their typing fluency-valence effect only in expert typists. The results reported here show the opposite: Experts had the weakest effects, and novices the strongest.

In short, there is no sense in which the typing speed data support the authors’ claim about expertise, fluency, and word valence. If they were interpretable, they would argue in the opposite direction.

**Analysis of Length and Frequency**

Although the authors clearly stated that word length and letter frequency would be used as control variables (as they were in J&C’s analyses), they then proceeded to treat them as predictors, and to interpret their main effects as evidence for “embodied” effects of typing on valence. Lots of problems here.

First of all, a variable shouldn’t switch from being a control to being a factor of interest post-hoc. Second, the effects of these variables are inconsistent across studies. Overall, however, there appears to be an effect of frequency on valence. This is unsurprising: both length and frequency are known to affect likeability – that’s why J&C controlled for them. But there is no justification for interpreting any effects of length or frequency on word valence as linked to *typing*. These effects might have nothing to do with typing; they could be mediated by ease/frequency of *reading*, ease/frequency of *pronouncing*, etc.

**2. The effect of RSA is reliable but not novel.**

Overall, there are no reliable effects other than the effect of RSA on Valence: The original QWERTY effect. All 9 of the analyses show effects of RSA in the predicted direction, controlling for Word Length and Letter Frequency (and sometimes other variables). 6 of these effects were significant; 2 were marginally significant as analyzed (in ANEW and the Dodds ‘Social Network’ corpus – however, J&C have already reported that the QWERTY effect in both of these corpora is significant when Word Length, Letter Frequency, *and their interaction* are controlled – which is how we decided *a priori* to analyze our data). One of the effects trended in the predicted direction but did not approach significance: this was in the Verb corpus, which was too small to be meaningfully analyzed, and did not show any significant effects -- of any variable.

Are these QWERTY effect replications reportable? Not really – certainly not in JML. Let’s go through them.

* The analyses in ANEW and AFINN cannot be reported: all of the significant (and nonsignificant) effects have already been reported previously by J&C.
* The Dodds Social Networking corpus only showed a marginal effect of RSA as analyzed, and J&C have already done an analysis of this corpus showing that the effect of RSA is significant when analyzed like our other corpora. The authors are aware of this analysis, which I pointed out to them in my last review (and they cited the source in this paper), yet they chose to present their analysis of this corpus as novel.
* The Verb corpus showed no significant effects.
* The combination of the 5 existing corpora is uninterpretable.
* Expt. 2 is uninterpretable.
* This leaves only one corpus for which there are any potentially reportable results: the ANEW extension. Although it’s reassuring to see the original QWERTY effect replicated in a much bigger sample of words, I don’t believe this analysis provides any new insights that would allow it to stand alone as a paper. First, it’s worth noting that Expanded ANEW encompasses Original ANEW, so this analysis is not independent of our original English ANEW analysis. Second, the authors might be tempted to try to single out this corpus in order to make a claim about a positive effect of Hand Alternation. They did show a significant main effect of Switch in this corpus, and there’s nothing uninterpretable about this effect, *per se*. However, interpreting this main effect would be a \*very bad idea\* considering everything else we know. J&C failed to find the Switch effect 5 times. The present authors failed to find the Switch effect 6 times. The fact that this effect appeared in the Expanded ANEW corpus does not erase the fact that it’s been searched for and not found over and over, in this paper and in J&C’s. The final straw: The slope of the main effect of Switch is positive in Expanded ANEW, but it’s *negative* in original ANEW.

**Concerns about the theoretical issues**

Three overlapping concerns, which I’ll mention in the hope that the authors are still interested in giving some thought to the QWERTY effect, despite this discouraging feedback about these two (sets of) experiments.

**1a. Did J&C fail to see the possibility of “embodied” influences on word meaning?**

As I mentioned in my last review, I find the rhetoric in this paper puzzling. The suggestion (toned down from the last version) is that somehow J&C have failed to consider the possibility of embodied influences of the typing process on word valence (i.e., the effect of variables that increase typing fluency on valence ratings). Actually, we considered 5 different potential sources of typing fluency in our paper: The whole paper was about testing whether typing fluency influences word meanings, and taking the first steps in understanding HOW this might work.

We considered all of the potential sources of typing fluency mentioned by these authors *and more*! The authors mention potential effects of (i.) hand dominance, (ii.) hand alternations, and (iii.) typing expertise, all of which we discussed in J&C 2012 – and in addition we discussed potential effects of (iv.) finger repetition and of (v.) asymmetries built into the keyboard, *per se*. In the end, we concluded that the initial data supported the Keyboard Asymmetry account most strongly, but that the Hand Dominance account had not been ruled out.

**1b. “Not just the right hand”**

This paper’s title (and rhetoric throughout) reflects a misunderstanding of our claim about what gives rise to the QWERTY effect: “Not just the right hand…” We never said it was just the right hand. On the basis of our first 5 corpus analyses (and the limited samples of left-handers), we concluded that typing fluency probably gave rise to the QWERTY effect, but that the source of this fluency was still up in the air: It could be use of one’s dominant hand, or it could be use of the less crowded side of the keyboard. To make it clear that we DID NOT want to claim that hand dominance was driving the QWERTY effect we (re-) named our dependent variable the Right SIDE Advantage (RSA); not the Right HAND Advantage. The present authors used both terms, but labeled their variable the RHA – thus implicitly making a more specific claim about the locus of the QWERTY effect than we were willing to make in the 2012 paper.

**1c. The reported experiments do not test the body-specificity hypothesis.**

These experiments did not test whether the QWERTY effect is body specific (i.e., different between right- and left-handers). To do so, it would be necessary to run a sufficient number of left-handers as well as right-handers, and to include handedness as a factor in the analyses (see J&C 2012, expts 1 and 3). We agree that the RSA results reported here are broadly consistent with body-specificity insomuch as they replicate our finding that the raters, who were (presumably) mostly right-handers, showed a Right Side Advantage. But without testing lefties or performing some additional experimental manipulation, it is not possible to interpret these data as unequivocal support for body-specificity.

**Summary**

Unfortunately, there is nothing to report here, despite a great deal of work on the authors’ part. I hope this painfully long review will help the authors understand why. Most of what’s significant is not novel. Most of what the authors portrayed as novel is not significant, or not interpretable. The authors wish to provide a more nuanced account of the QWERTY effect by showing effects of other fluency-relevant variables like Switching and Typing Speed on word valence. Such an account would be welcome, and would be consistent with J&C’s proposal that manual motor fluency should lead to a link (or links) between typing and word valence. But the evidence for the effects of these variables is very, very weak, and some of the scant effects the authors are interpreting go the opposite direction from their predictions.

There are serious problems with the scholarship, the methods, the data, and with the interpretation of these data and of J&C’s claims. I feel some frustration about having to reiterate some of the most important points from my last review. However, I appreciate the authors’ interest in extending work on the QWERTY effect, and I’ll repeat the offer I made last time: Please feel free to contact me if you have any questions about this review, or to discuss productive ways to pursue this line of research.