Reviewer #2: This manuscript presents two studies. In the first study, the QWERTY effect (whereby words are rated as being more positive when they are typed mostly with the right hand) was replicated with a broader set of measures than used in the original work on this topic. The second study examined the role of expertise in these effects, and showed that expertise did interact with the other variables of interest.  
  
I enjoyed reading this paper. The topic is interesting and timely, and I think the authors are pursuing the phenomenon in an interesting way. Nonetheless, there are a couple of issues that need to be addressed before this paper can be considered suitable for JML.  
  
1 --  The paper is a bit too narrowly focused. The authors draw on frequency and fluency as explanatory constructs, and consider these to be signs of embodiment (e.g., fluency of motor programs, ease of action...). This is fair enough, but these constructs have been used broadly in the literature -- e.g., the mere exposure effect, Bruce Whittlesea's studies of learning and memory, and so on -- and are not always associated with embodiment. I think this work would benefit from a deeper consideration of the mechanisms at play in driving these effects.  
  
2 -- It is unfortunate that the results of Experiment 1 are a bit inconsistent. I think the authors should do a bit more thinking about why this is, rather than pointing to some differences in the norms and leaving it at that.  If we are to understand the effects at hand, it is important to get a handle on the conditions under which they will appear (or not).  
  
  
  
  
  
Reviewer #3: This is a replication and extension of the QWERTY effect - that words typed with letters from the right hand side of the keyboard are more positive in valence than those from the left hand side. In principle I am sympathetic to a simple replication. Following the publication of the original results in Psychonomic Bulletin and Review, there was quite a furor surrounding the reliability of the findings, as exemplified by the analyses conducted by Mark Liberman on Language Log. The present paper claims to have: "found that the right hand advantage is replicable to new valence stimuli, as well as experimental manipulation." However, I do not share the authors interpretation of their results, and do not think the paper is strong enough for publication in JML.  
  
To summarise my main objection: the Right Hand Advantage (sometimes abbreviated in the paper as RHA, other times as RSA) is neither robust nor reliable. After examining 5 databases the authors find a significant RHA in only 2 out of 5 databases. In a couple of other places the authors find an interaction between RHA and hand switching (i.e., switching between hands when typing a word), or between RHA, hand switching and third variable. But these interactions are also not consistent. In all cases, the variance account for is extremely small -- pr2 between .001 and .006. So, as a replication this paper is just not convincing. The effects are flakey. Perhaps this is consistent with what was reported in the original paper, but this certainly does not support the strong claims made in the abstract and conclusions of this paper.  
  
Some other issues:  
Why is frequency of the words not taken into consideration? The authors control for letter frequency but surely word frequency is also likely to be a confounding variable (the mere exposure effect).  
  
Typing manuals were consulted and letters from the keyboard was assigned to left and right hands as follows:  
Q  W  E  R  T  ///  Y  U  I  O  P  
A  S  D  F  G  ///   H  J  K  L  
Z  X  C  V  B  ///  N  M  
According to this assignment, 15 letters are typed with the left hand and only 11 are typed with the right hand. I'm just not convinced that this is how people actually type these keys. Do naïve participants, in fact, type according to this prescriptive guide? That seems to be a crucial assumption of this study, and I do not see the relevant evidence in support.  
  
Valence rankings from the databases differed and were rescaled to a 1-9 scale. How was this rescaling achieved? Why not use the original scores?