8-23-2017

Individual Differences in Sensitivity to Homophony in Visual Word Recognition

Henry George Wolf VII
henry.wolf@uconn.edu

Individual Differences in Sensitivity to Homophony in Visual Word Recognition

Henry George Wolf VII

M.A. Teaching English to Speakers of Other Languages, Columbia University, 2011

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

At the

University of Connecticut

2017

APPROVAL PAGE

Master of Science Thesis

Individual Differences in Sensitivity to Homophony in Visual Word Recognition

Presented by

Henry George Wolf VII, M.A.

Major Advisor ____

Jay Rueckl

Associate Advisor ____

Gerry Altmann

Associate Advisor ____

Nicole Landi

Associate Advisor ____

James S. Magnuson

University of Connecticut

2017

Acknowledgements

The researcher wishes to express his sincere gratitude and appreciation to the following individuals, who contributed a great deal in helping shape this paper, provided support throughout the process, or both.

Dr. Jay Rueckl, thesis advisor, for providing guidance and patience during this project, from conceptualization through the countless revisions of the manuscript. This has increased both my writing skills and my ability to think scientifically.

Drs. Gerry Altmann, Nicole Landi, and Jim Magnuson, committee members, for their consideration of this manuscript and evaluation of my defense.

Dr. Julie Brown, lab manager, for scheduling and managing the research assistants during collection of data, as well as for scoring tasks.

Tong Li and Oliver Sawi, lab mates, for leading the way and providing excellent examples of how masters theses (and general examinations) are to be written and defended.

Undergraduate research assistants, for dutifully running the experiment, collecting and scoring much of the data.

Table of Contents

Abstract

Most research involving division of labor and visual word recognition has focused on the typical reader. More recently, there has been a shift toward research involving individual differences in division of labor between readers. While the imageability effect has been established as a measure of individual differences in use of the semantic pathway, a measure of the phonological pathway has yet to be established. The current study investigated the homophone effect in a semantic categorization task as one such possible measure. Data was also collected regarding imageability, wordlikeness, and pseudohomophony in a lexical decision task. Additionally, participants completed a battery of ID measures as a more holistic measure of performance. Each of the main effects replicated the results of the previous literature. Participants were found to differ in individual variability, however there was less variability in the homophone effect in RT. In general, participants with larger effects tended to make fewer errors and respond more slowly. Further, I found patterns of relationships between the ID battery and the effects in RT, but not those in error rate. The data suggests that RT may capture individual differences better than error rate and that the lexical decision task may capture individual differences better than the semantic categorization task.

Keywords: reading, visual word recognition, individual differences, division of labor, wordlikeness, imageability, homophony, pseudohomophony

Individual Differences in Sensitivity to Homophony in Visual Word Recognition

Introduction

Visual word recognition involves accessing phonological and semantic information from orthographic information. Numerous factors can affect the speed and ability of readers to access the correct information. As an example, word frequency been shown to affect responses to naming and lexical decision tasks. Studies have shown that high frequency words are read more quickly and accurately than low frequency words (Forster & Chambers, 1973). Regularity, which has to do with whether the relationships between spelling and sound in a word are those most common, is another factor shown to affect speed and accuracy of responses in naming and lexical decision tasks (Stanovich and Bauer, 1978). Certain factors, such as these, have also been shown to interact (Andrews, 1982), suggesting a complex relationship between multiple factors is involved in word reading. These are just a two of the factors that have been identified in the word recognition literature.

While most research has centered around the typical reader, there is clear evidence for individual differences between readers in visual word recognition. Studies that have investigated individual differences between readers in word recognition have provided insights into what is shared versus variable. For example, Jorm (1977) found that while high frequency words were easier to read aloud for both good and poor readers, high imageability words were easier to read aloud only for poor readers. Butler and Hains (1979) found that participants with larger vocabularies showed a smaller effect of word length than those with smaller vocabularies. They also found that participants with larger vocabularies responded faster to a naming task than those with smaller vocabularies, but responded more slowly to a lexical decision task, indicating that the degree to which individual differences can be captured may be influenced by task demands.

Research examining individual differences between readers has gained more attention in recent years. Yap, Balota, Sibley, and Ratcliff (2012) found that higher vocabulary knowledge is associated with faster and more accurate visual word recognition. Differences between good and poor readers have also been found with regards to the way morphemic information (Kuperman & Van Dyke, 2007) and attention (Herdman & LeFevre, 1992) contribute to word recognition. Each of these results suggests the existence of important differences in how individuals read.

Division of Labor

Considering the behavioral evidence for individual differences between readers, it seems likely that there would be individual differences in the manner they read. Theories of word reading describe the ways people might use different pathways to read (see Coltheart, Curtis, Atkins, & Haller, 1993). There are at least two pathways that could be used to read most words aloud. One pathway would map the orthographic information to the phonological information while another would first access semantic information, which could then be used to access the phonological information. Consistent words (e.g. hint) could be read correctly via either pathway, but the semantic pathway may be better suited for reading inconsistent words (e.g. pint). Conversely, the phonological pathway is better suited to reading nonwords. These pathways have also been inferred through behavioral effects and the results of neuroimaging studies (e.g. Hoffman, Lambon Ralph, & Woollams, 2015; Newman & Joanisse, 2011). Both pathways are needed to explain such results, but the degree to which an individual relies on one pathway or the other may differ systematically.

Early descriptions of the dual route model suggested that the phonological pathway involved the application of grapheme-phoneme (spelling-sound) correspondence rules while the semantic pathway involved a direct dictionary lookup of the pronunciation of a word (Forster &

Chambers, 1973). These models accounted for frequency effects by stating that the look-up method is faster than the application of rules and that more frequent words would be found more quickly. In this framework, most words would be read through the lexical pathway, with the phonological pathway being used primarily for low frequency words and nonwords, as well as during learning to read (Coltheart, Curtis, Atkins, & Haller, 1993). Consistency effects were thought to arise from the interaction between the two pathways, with the relationship between frequency and consistency being due to rules being applied only after the dictionary lookup process (Stanovich & Bauer, 1978). The predictions arising from the dual route theory have informed many research studies examining division of labor.

Baron and Strawson (1976) performed some of the earliest research on individual differences in division of labor in mapping orthography to phonology. They sought participants who were skilled in only one pathway because they expected them to show different effects of consistency in a naming task. Participants primarily using the semantic pathway would be expected to show smaller consistency effects than those primarily using the phonological pathway. They used measures of spelling and nonword reading as indicators of skill in using the two pathways. They called participants who showed strong knowledge of the spelling-sound regularities, as measured by the nonword reading task, but weak knowledge of spelling, as measured by the spelling recognition task, "Phonecian", and those that showed the opposite pattern "Chinese". As they predicted, the Phonecian readers showed a larger consistency effect in the naming task than the Chinese readers, suggesting the former gave preference to the phonological pathway while the latter gave preference to the semantic pathway.

The assumptions made by Baron and Strawson in determining which participants fell into each category were criticized by Brown, Lupker, and Colombo (1994). In addition to citing a

lack of control over word frequency and the use of word lists, rather than individual words, in the naming task, they state that Baron and Strawson did not sufficiently establish that the, renamed, “Phoenician” and “Chinese” readers performed better in tasks where their respective phonological and semantic pathways should be more beneficial. They attempted to replicate the results of Baron and Strawson with better controlled tasks, including an extended oral spelling task and multiple naming tasks. Were the difference between these two groups qualitative, the Phoenician readers should have shown a smaller lexicality effect, as more words would be read via their phonological pathways. Meanwhile the Chinese readers would have shown smaller frequency and consistency effects. The results showed the reverse, however, with the Chinese readers showing larger frequency and consistency effects.

More recent research regarding division of labor in reading has been influenced by connectionist theories, such as the Triangle Model (Seidenberg & McClelland, 1989). These are learning models which are used to explain the development of reading and division of labor over time by computing the relationships between orthography, semantics, and phonology without predetermined rules. In this type of neurally-inspired model, a word is presented which generates a pattern of activation that is mapped to semantics or phonology. When the meaning or pronunciation is incorrect, the mappings are modified, affecting both accuracy and division of labor between the pathways. The relative use of each pathway is largely determined by the structure of the mappings between the orthographic, phonological, and semantic representations, as well as factors such as frequency, consistency, and experience. Unlike dual route models, connectionist models suggest that both the semantic and phonological pathways contribute to the production of each word in parallel, even among skilled readers (Harm & Seidenberg, 2004). Frequency effects can be accounted for because the learning process results in stronger mappings

for high than low frequency words. Likewise, nonwords will not have been seen at all, but sub-lexical information makes them pronounceable through the phonological pathway. Due to the use of both pathways, these models are also capable of capturing a wider range of psycholinguistic effects. Connectionist theories have greatly influenced recent research in word reading and individual differences.

Recently, there has been heightened interest in individual differences in the semantic pathway to phonology. Perhaps the most widely studied semantic effect is imageability. The Triangle Model predicts that semantic effects in word naming would be greatest when naming low frequency inconsistent words because they would be improperly named via the phonological pathway and provide fewer opportunities for learning. Strain, Patterson, and Seidenberg (1995) conducted a series of three naming experiments in which they manipulated word frequency, consistency, and imageability. They collected response times and error rates for normal and speeded naming tasks, and found a clear three-way interaction. In general, participants produced more errors and responded more slowly to low versus high imageability words, but only when they were low frequency and inconsistent. This suggests that the phonological pathway is more efficient for high frequency and consistent words.

Strain and Herdman (1999) followed this study by examining individual differences in the imageability effect in word naming. As a measure of ability with the phonological pathway, they first categorized participants by phonological skill using the Word Attack and Sound Blending tasks from the Woodcock-Johnson reading test. Participants then performed a naming task with the imageability and consistency of words manipulated. Participants with lower phonological skill showed larger imageability effects, indicating greater use of the semantic pathway. While larger consistency effects may be expected, due to poorer phonological skill,

these participants also showed smaller consistency effects in RT, possibly indicating less use of the phonological pathway. However, the expected pattern was seen in error rates, where a larger effect was shown. The relationship between the semantic contribution and imageability was confirmed by Woollams, Lambon Ralph, Madrid, and Patterson (2016) in an individual differences study of naming.

As the imageability effect in word naming has been identified as a useful measure of individual differences in the semantic pathway during naming, one might expect there would be similar measure of the phonological pathway to meaning. However, such an effect has yet to be firmly established. The current study examines the viability of the homophone effect in semantic categorization as a candidate, as it provides evidence a shared phonological representation can affect the meaning accessed. Further, it could prove capable of capturing variation with regards to division of labor in visual word recognition.

The homophone effect was investigated by Van Orden (1987) using a series of semantic categorization tasks. Subjects were presented with a category name above a fixation point, which was followed by a target word, which was then replaced by a pattern mask. They then had to indicate whether the target word was a member of the category. False positive error rates were significantly higher for the homophones than the controls. In addition, the error rates were higher for similarly spelled homophones and controls than for those less similarly spelled, indicating an interaction between the effects of homophony and orthographic neighborhood. Van Orden suggested that these results support reading models involving the use of both phonological and orthographic information in word identification. In another experiment Van Orden manipulated exemplar and target word frequency, finding more errors for low frequency exemplars but no effect of frequency among the target stimuli.

Jared and Seidenberg (1991) were critical of the Van Orden results due to the use of narrow categories (e.g. “part of a horse's harness”), which may have resulted in the phonological representations of category exemplars being primed before stimulus onset. They performed a replication with a larger set of target words and spelling controls, as well as broader category names (i.e. “living thing” and “object”). With these broader categories, the homophone effect was found to be significant only for homophones with low frequency exemplars. While this suggests that the categories used by Van Orden may have primed phonological representations of the exemplars, resulting in an increased error rate, the existence of the effect was further established under certain circumstances.

Few studies have yet examined individual differences in the homophone effect and its relationship to other behavioral measures. Lewellen, Goldinger, Pisoni, and Greene (1993) used a battery of nine measures (i.e. word familiarity ratings, the Nelson-Denny Vocabulary Test, Author and Magazine Recognition Tests, a spelling test, a language experience questionnaire, and verbal and math SAT scores) to differentiate between subjects with high and low skill on these tasks. In general, those with high skill had shorter reaction times and smaller error rates than those with low skill, suggesting greater reliance on the semantic pathway, while those with low skill showed greater reliance on the phonological pathway. Jared, Levy, and Rayner (1999) used a different battery of measures to differentiate between good and poor readers. They examined the homophone effects using proofreading and eye tracking tasks and came to the same conclusion, that while good readers are more efficient with both pathways, they primarily make use of the semantic pathway.

Indicators of Division of Labor

By examining the ways in which behavioral effects are related, it may be possible to better identify useful indicators of division of labor. The homophone effect alone cannot provide a complete description of the division of labor for a reader, as there are multiple possible reasons a reader could show a large homophone effect. The participant may rely upon the phonological pathway due to an inefficient semantic pathway or a highly efficient phonological pathway. Alternatively, as noted by Starr and Fleming (2001), it may be that the reader is poor at spelling or has had especially limited experience reading low frequency lexical items. Interactions with other behavioral effects of reading, as well as measures of individual differences, may provide a way of differentiating between these possibilities.

Imageability. Influences that are primarily semantic in nature should affect both pathways in a lexical decision task, as the degree to which there is target semantic information should be helpful in determining whether a character string is a word, especially when no word-formation rules have been broken among nonwords. It should also be beneficial in a semantic categorization task because the semantic information about a word is, perhaps, the most important piece of information when determining whether a word belongs to a category. As readers of all ages have a great deal more experience with the mappings between phonology and semantics than with those between orthography and semantics, one might expect targets with less semantic information to be accessed more easily via the phonological route. Put another way, they might be more disruptive of the orthography to semantics pathway. The imageability effect is the result of differences in ability to visualize a word. More imageable words (e.g. "shovel") result in shorter response times than less imageable (e.g. "swift") in lexical decision. Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) investigated several semantic effects,

including imageability, and found reliable effects in a lexical decision task. Combining information about the size of the imageability effect with that of the homophone effect, one might suggest that a reader with a large homophone effect, which is primarily phonological, and a small imageability effect primarily uses the phonological route. If a participant has a small homophone effect and a large imageability effect, one might suggest this reader is relatively more skilled with the semantic pathway. Without both measures, it is difficult to make such a comparison.

Wordlikeness. Influences that are primarily orthographic in nature should also affect both routes to meaning because orthography is the starting point for both pathways. As the mappings from orthography to phonology in English are less complex than mappings directly from orthography to semantics, one might expect orthographic influences to have a greater influence on the denser phonological pathway. Wordlikeness is related to the mapping between orthography and phonology. BAME is more wordlike because there are many words that are similar (e.g. GAME and BANE), whereas YNZX is less wordlike because multiple letters would have to be changed to make a real English word. Coltheart, Davelaar, Jonasson, and Besner (1977) showed that more wordlike nonwords are responded to more slowly than less wordlike nonwords in a lexical decision task. Research by Yap, Balota, Cortese, and Watson (2006) showed a continuum that spanned less wordlike nonwords, more wordlike nonwords, and pseudo-homophones in lexical decision with increasing response times. One might expect the wordlikeness and homophone effects to align, because of the less complex relationship between orthography and phonology. However, if a participant has poor spelling knowledge, one might expect a large homophone effect and a small wordlikeness effect. Conversely, if a participant has better spelling, the results may show a small homophone effect and a large wordlikeness effect.

Pseudohomophony. Influences that are related to the mapping between phonology and semantics in nature should affect primarily the phonological route to meaning, because they are not a part of the semantic route. These effects can provide the clearest evidence for individual differences in division of labor in mapping from orthography to semantics. As these measures can arise from different tasks, they can provide some evidence of stability across tasks and measures. The homophone and pseudo-homophone effects arise from stimuli that have differing mappings from orthography to phonology, but similar mappings from phonology to semantics. Pseudo-homophones (e.g. BAIR) are nonwords orthographically but can be read as words phonologically. As homophones (e.g. BARE) are words both orthographically and phonologically, the main difference between these two types of stimuli is lexicality. Studies of children have shown that the pseudo-homophone effect shrinks as reading ability increases (Grainger, Lete, Bertrand, Dufau, & Zeigler, 2012). Participants with poorer spelling may show effects of homophony and pseudohomophony that are similar, as they may not know the correct spelling of the target word. Both effects should be smaller for those with better spelling, but the homophone effect may be larger as stimuli are all correctly spelled lexical items while the pseudo-homophonic stimuli are not.

Aims and Predictions

I aim to investigate the use of homophony as a marker of individual differences in use of the phonological pathway in accessing word meaning through two experimental tasks and a battery of more holistic individual differences tasks (henceforth “ID battery”). A semantic categorization task provides the basis for our investigation of the homophone effect while a lexical decision task allows for an examination of imageability, wordlikeness, and pseudo-homophony. The ID battery consisted of the Author Recognition Task (Cunningham, A. E., &

Stanovich, K. E., 1990), the Nelson-Denny Vocabulary Test (Nelson, M. J., Brown, J. I., & Denny, M. J., 1960), a spelling test, and the pseudoword and sight word reading efficiency tasks from the Test of Word Reading Efficiency (Torgesen, J. K., Wagner, R. K., & Rashotte, C. A., 2012). These ID tasks were used to get an understanding of skills related to reading and to examine relationships with the experimental tasks.

I predict that individual differences in the homophone effect will systematically vary with individual differences in the imageability, wordlikeness, and pseudo-homophone effects. I might also expect the effects measured in RT to show a positive relationship with those in error rates. Additionally, participants with larger effects are expected to be slower and produce more errors. Better performance on the ID battery should correlate with success on the measures in the experimental tasks. Further, I expect that participants who perform better on the tasks in the ID battery will have smaller effects in these tasks. I also suspect there will be a relationship between the size of the behavioral effects and these ID measures that is similar in kind of the relationship between the effects and experimental tasks. Most importantly, I predict the homophone effect will prove itself a useful marker of phonological pathway use.

Method

Participants completed two experimental tasks and an ID battery during a single session. Most measures in the ID battery were performed in a web browser using Qualtrics while the experimental tasks were performed in e-Prime on a separate computer that was not connected to the Internet. All instructions were presented on screen and read aloud by a proctor. Participants were seated at the web-connected computer and first filled in demographic information. They then completed the Author Recognition Test (ART) before switching computers and performing the SCT. Subtests of the Test of Word Reading Efficiency (TOWRE-2) were recorded using an

audio recorder while a participant remained seated. Participants then completed the LDT. They switched computers once again to perform the spelling and vocabulary tests on Qualtrics. This order was used in order to provide participants with breaks between tasks. Feedback was not provided for any of the tasks. At the end of the session, participants were debriefed on the purpose of the study.

Participants

Participants included 118 undergraduates at the University of Connecticut who received course credit for their participation. Due to computer error, eight participants were removed from analysis. One additional participant was removed, as this participant was not a native speaker of English. Another was removed due to performing near chance on word items in the LDT, indicating an inability to complete the task properly. Of the 108 remaining participants, 77 were women and 31 were men. All were native English speakers and had normal or corrected-to-normal vision.

Semantic categorization

The first experimental task was designed to examine the effect of homophony on RT and error rate.

Design and materials. Each participant saw 174 target words, split among three categories. Participants saw each target word only once. The practice trials included a total of 15 “no” fillers and 15 “yes” fillers. To avoid the induction of specialized processing strategies, fewer than 17% of stimuli were homophones. The experimental trials included a total of 24 homophones (e.g. TOE for TOW), 24 spelling controls (e.g. TON for TOW), 24 “no” fillers, and 72 “yes” fillers. As such, there was an equal number of category members, words that belong to the presented category, and nonmembers.

Stimuli consisted primarily of words collected for this experiment, however 16 were adapted from Jared and Seidenberg (1991, p. 391) and eight were adapted from Pexman, Lupker, and Jared (2001, p. 155). All homophones and spelling controls were matched with low frequency category exemplars; fillers were not matched. Orthographic similarity (OS) was calculated in the same manner as was done by Van Orden (1987, p. 196), which was adapted from the graphic similarity measure devised by Weber (1970). Mean OS, in relation to the category exemplars, was 0.64 for the homophones and 0.62 between for the spelling controls. Mean log frequencies from the Hyperspace Analogue to Language frequency norms (Lund & Burgess, 1996) of approximately 131 million words were gathered from the English Lexicon Project (Balota et al., 2007). Frequencies for category exemplars (7.92), homophone foils (7.20), and spelling controls (7.82) were limited to low frequency words. Mean stimulus length was also controlled for as much as possible with regards to the category exemplars (4.33), homophone foils (4.71), and spelling controls (4.42).

Procedure. Participants first performed the 30 practice trials. After confirming that they understood the task, participants completed the 144 experimental trials, with a break available after every 48 experimental stimuli. Presentation order was randomized for each subject. The task required approximately 25 minutes to complete.

During each trial, participant saw a fixation point (“+”) in the center of the screen for 250 ms followed by a category name (e.g. “living thing”, “object”, or “action”) for 2,000 ms and then a target word for 250 ms. A pattern mask (“+++++++”) was then presented that lasted until the participant responded “yes, it belongs to the category” by pressing “1” or “no, it does not belong to the category” by pressing “0” on a keyboard. Participants were provided a maximum of 3,000 ms to respond. There was a pause of 1,500 ms between the removal of the pattern mask and the

presentation of the next fixation point. All instructions and stimuli were presented in green capital letters in the Arial font on a black background.

Lexical decision

The second experimental task was designed to examine the effects of wordlikeness, pseudohomophony, and imageability on RT and error rates.

Design and Materials. Each participant saw 270 target stimuli. 30 in the practice and 240 in the experimental trials. Participants saw each target stimulus only once. The practice trials included 15 “no” fillers and 15 “yes” fillers. Nonword stimuli were divided evenly by wordlikeness and included 60 pseudohomophones (e.g. CAIK, CHACE) and 60 pseudowords (e.g. CHYZE, CLEEP). Experimental word stimuli were divided evenly by imageability. An additional 40 original “yes” fillers were included to have an equal number of words and nonwords.

Nonword stimuli were adapted from Pexman, Lupker, and Jared (2001, p. 156) while experimental word stimuli were adapted from Evans, Lambon Ralph, and Woollams (2012). Frequency data was gathered in the same manner as for the SCT. All pseudohomophones and pseudowords were matched with low frequency words. Frequencies for matched words (8.65) were limited to low frequency words as much as possible. This is also true of the mean experimental word frequencies (8.50). Mean nonword length was also controlled for with regards to the matched words (4.72), pseudohomophones (4.63), and pseudowords (4.63). Mean word length (4.00) was slightly lower than for the nonwords.

Procedure. Participants first performed 30 practice trials. After confirming that they understood the task, participants completed the 240 experimental trials, with a break available

after every 60 experimental stimuli. Presentation order was randomized for each subject. The task required approximately 10 minutes to complete.

During each trial, participants saw a fixation point (“+”) for 500 ms, which was followed by a target stimulus. The target stimulus remained visible until the participant responded “yes, it is a word” by pressing “1” or “no, it is not a word” by pressing “0” on a keyboard. Participants were provided a maximum of 3,000 ms to respond. There was a pause of 500 ms between the removal of the target word and the presentation of the next fixation point. Colors and fonts during presentation were the same as in the SCT.

Individual differences battery

The measures selected for inclusion in the ID battery obtained information on participants’ reading experiences and abilities.

Demographics. Demographic data was collected but not included in the correlations below. In addition to gender, race, and ethnicity, questions were posed to ascertain the language background of each participant.

Author Recognition Task (ART). Participants were presented with a list of 66 names, half of which were names of authors, and were instructed to click "Yes" if the name belonged to an author and "No" if it did not (modified from Cunningham, A. E., & Stanovich, K. E., 1990). In order to use a single metric that included both speed and accuracy, inverse efficiency was calculated by dividing the amount of time required to complete the task by the number of correct responses, resulting in a lower efficiency score indicating higher performance.

Nelson-Denny Vocabulary Test. Participants were presented with sentences with one of the words missing and five multiple choice word options. Participants were instructed to click on the word that accurately completed the sentence (modified from Nelson, M. J., Brown, J. I., &

Denny, M. J., 1960). Items were of increasing difficulty. Efficiency was calculated in the same manner as for the ART.

Spelling Recognition Task. Participants were presented with 80 pairs of strings. One member of each pair was a word spelled correctly while the other was the same word misspelled. Efficiency was calculated in the same manner as for the ART.

Test of Word Reading Efficiency (TOWRE-2). Participants were first presented with a list of sight words of increasing length and complexity (Torgesen, J. K., Wagner, R. K., & Rashotte, C. A., 2012). Participants read the words aloud and in order, as quickly and accurately as possible. Responses were recorded until the end of a 45 s time limit. This process was repeated for a list of pseudowords. The pseudoword and sight word reading tasks were conducted and recorded offline, with two raters scoring each of the recordings independently. In the case of disagreement between raters, a third rater acted as arbitrator. Separate inverse efficiency scores were calculated for the sight words and pseudo-words by dividing the time used in reading from each list by the number of acceptable pronunciations, resulting in a lower inverse efficiency score indicating higher performance.

Results

Analysis of the data began with an examination of group-level task success and effect sizes. This was followed by an investigation of the variability in individual differences between participants. Relationships among the experimental tasks were then considered to investigate whether RT is related to error rate, whether both are related to effect size, and whether the effects are related. Finally, I examined relationships between and within the ID battery and the experimental tasks. This was done to determine whether the ID measures are related to one another, the experimental tasks, and effect sizes. The remainder of this section is organized with

this structure of results in mind. Specific effects and their relationship to other metrics will be presented. All analyses were performed using R (R Core Team, 2016).

Experimental task analyses. Analysis of the data began with an examination of the RT and error rate means of means for each experimental manipulation. RTs were log transformed to limit the effect of very long responses. Error rates were converted to log odds (logits) due to the binary nature of this metric. Mixed effects regression models were conducted separately for word and nonword items from lexical decision task, as well as for the items from the semantic categorization task. A total of six models were used to analyze the RT and error rate data from the experimental tasks using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Models with RT as the dependent variable of interest were calculated using linear mixed effects regression (LMER) whereas those with error rate as the dependent variable were calculated using generalized linear mixed effects regression (GLMER) models, which is more appropriate for the categorical nature of each response (Jaeger, 2008). The independent variables of interest were included using deviation coding (-0.5, 0.5). This was done progressively, first adding the fixed then random effects, so that models could be compared. Items were controlled for by entering them as random factors to prevent individual variance from items from affecting the experimental effects. Trial number and previous RT were standardized and controlled for to prevent effects of the random presentation order and preceding items from affecting the current item.

Group-level task analyses. Each of the group-level effects was as expected based upon previous research. The means of means in RT and error rate by experimental manipulation were all in the expected directions (Table 1). These effects were shown to be significant in all cases. Participants responded more slowly (b = 0.073, $X^2(1) = 6.9$, p < .01) and made more errors (b =

1.088, $X^2(1) = 10.9$, $p < .001$) on homophones than spelling controls. They also responded more slowly (b = 0.089, $X^2(1) = 18.3$, $p < .001$) and made more errors (b = 1.417, $X^2(1) = 17.7$, $p < .001$) on low than high imageability words. In addition, they responded more slowly (b = 0.052, $X^2(1) = 11.6$, $p < .001$) and made more errors (b = 0.890, $X^2(1) = 18.7$, $p < .001$) on pseudo-homophonic nonwords than non-homophonic nonwords. Further, they responded more slowly (b = 0.079, $X^2(1) = 28.3$, $p < .001$) and made more errors (b = 0.897, $X^2(1) = 19.0$, $p < .001$) on more than less wordlike nonwords. Table 2 describes the models from which these results were derived.

Individual differences analyses. By participant individual differences were added to each model. All were found to be significant (Table 2). I compared the participant means to their estimates and effects as captured by the models (Table 3) in order to ensure the model estimates were a good measure of ability to complete the task and effect size. There were very strong relationships between the participant means and by participant deviation from the group intercept. Relationships for the semantic categorization task are shown in the top half of Figure 1. There were also strong relationships between most of the effects as measured by difference scores and as captured by the models. The two notable exceptions were between the difference scores and the model predicted individual effects of wordlikeness, $r(106) = .36$, $p < .01$, and imageability, $r(106) = .08$, *ns*. These relationships are shown in the bottom half of Figure 1. As all estimates and most effects showed strong relationships, they are used throughout the rest of the results (see Baayen, Davidson, & Bates, 2008).

Histograms were produced to examine variability in effect sizes in RT (Figure 2) and error rate (Figure 3). With regards to RT, the effect with the widest variability was pseudohomophony, SD = 0.032. Variability in the effects of wordlikeness (SD = 0.024) and

imageability (SD = 0.023) were quite similar, while that in the homophone effect was rather limited, SD = 0.019. In the error rates, the widest variability was found in the imageability effect, SD = 0.706. Variability in the pseudohomophone effect was also large, SD = 0.541. Variability was smaller in the homophone effect (SD = 0.349) and the wordlikeness effect (SD = 0.166). Having established that mixed effects models are capturing variability between participants and that they are closely related to the raw participant data, the random intercepts and slopes from the models were used in as measures of average task success and effect size in the following correlations (ibid).

Relationships among experimental tasks. The top-left quadrant of Table 4 shows correlations within the experimental tasks. Within these measures, I found that participants who are faster on one task tend to be faster on the other as well. Participants respond faster to the words from the lexical decision task tend to also respond faster to the nonwords, $r(106) = .75$, $p < .01$ (Figure 4, left). Even across tasks, faster participants in one tended to be faster in the other, with moderate correlations between the semantic categorization task and the words, $r(106) = .45$, $p < .01$, as well as the nonwords, $r(106) = .53$, $p < .01$, from the lexical decision task.

I then considered relationships among error rates and found that participants who perform better on one task tend to perform better on the other, but these relationships were not nearly as strong. I found weak positive correlations between the intercepts from the lexical decision nonwords and words, $r(106) = .22$, $p < .05$, and semantic categorization, $r(106) = .33$, $p < .01$ (Figure 4, right). Further, the relationship between error rate on the lexical decision words and semantic categorization was not significant, $r(106) = .06$, *ns*.

Finally, I examined relationships between both measures in our experimental tasks to determine whether there is a relationship between speed and accuracy. Our results do not show a

tradeoff between speed and accuracy. Weak positive correlations were found between the RT and ER intercepts from the semantic categorization task, $r(106) = .26$, $p < .01$, and the lexical decision nonwords, $r(106) = .24$, $p < .05$, while no relationship was found among the lexical decision words, $r(106) = .01$, *ns*.

Relationships between experimental tasks and effects. The bottom-left quadrant of Table 4 shows correlations between the experimental tasks and effects. Relationships between tasks and effects were examined to determine whether effect size is related to task success. In general, participants with larger effects also responded more slowly to the task. Participants with larger homophone effects were slower to respond to the semantic categorization task, $r(106) = .30$, $p < .01$, but not to the lexical decision task. Those with larger imageability effects were slower to respond to the words, $r(106) = .75$, $p < .01$, as and nonwords, $r(106) = .54$, $p < .01$, from the lexical decision task, in addition to the semantic categorization task, $r(106) = .34$, $p < .01$. Participants with larger pseudohomophone effects responded more slowly to the nonwords, $r(106) = .59$, $p < .01$, and words, $r(106) = .54$, $p < .01$, in the lexical decision task, as well as to the semantic categorization task, $r(106) = .31$, $p < .01$. Similarly, participants with larger wordlikeness effects responded more slowly to the nonwords, $r(106) = .59$, $p < .01$, and words, $r(106) = .45$, $p < .01$, in the lexical decision task, and to the semantic categorization task, $r(106) = .31$, $p < .01$. The relationships between each effect and task in RT can be seen in Figure 5.

While there was a pattern of larger effects being related to smaller error rates, fewer significant correlations were found in this measure than in RT. Participants with larger homophone effects made fewer errors only in semantic categorization, $r(106) = -.29$, $p < .01$. The relationship between the imageability effect and the words from the lexical decision task was much stronger, $r(106) = -.91$, $p < .01$. Additionally, participants with larger imageability effects

made fewer errors on the nonwords from the lexical decision task, $r(106) = -.24$, $p < .05$. Also, participants with larger wordlikeness effects made fewer errors on the words from the lexical decision task, $r(106) = -.25$, $p < .05$. No significant relationships were found between task the effect of pseudohomophony in error rates. The relationships between each effect and task in error rate can be seen in Figure 6.

Relationships among the experimental effects. The bottom-right quadrant of Table 4 shows correlations within the experimental effects. In general, participants with larger effects in RT are trending toward larger effects in error rate. This is clearest in homophony, $r(106) = .23$, $p < .05$, but also significant in imageability, $r(106) = .16$, $p < .05$. Correlations across measures for neither pseudohomophony nor wordlikeness were significant, however, making it difficult to suggest the measures are equally capable of capturing individual differences.

Relationships within the experimental effects were then investigated for indications of division of labor. Participants with larger effects in one manipulation tended to have larger effects of another manipulation, with exceptions being between the effect of wordlikeness and the effects of homophony and pseudohomophony in error rates. Participants with larger effects of imageability often also had larger effects of pseudohomophony, $r(106) = .47$, $p < .01$, and wordlikeness, $r(106) = .48$, $p < .01$. Even more so, participants with larger effects of wordlikeness also had larger effects of pseudohomophony, $r(106) = .78$, $p < .01$. In error rates, participants with larger effect of homophony also had larger effects of imageability, $r(106) = .16$, $p < .05$, and pseudohomophony, $r(106) = .24$, $p < .05$. Interestingly, the only significant negative correlations were found between the effect of wordlikeness and the effects of homophony, $r(106) = -.21$, $p < .05$, and pseudohomophony, $r(106) = -.79$, $p < .01$, in error rate (Figure 7).

Finally, it should be noted that there was a general pattern of positive correlations across measures for effects involving words but not nonwords. While the homophone effect in error rate was not strongly related to the imageability effect in RT, $r(106) = .13$, *ns*, there were significant correlations with the effects of pseudohomophony, $r(106) = .26$, $p < .01$, and wordlikeness, $r(106) = .25$, $p < .05$, in RT. The imageability effect in error rate was related to the effects of homophony, $r(106) = .24$, $p < .05$, pseudohomophony, $r(106) = .18$, $p < .05$, and wordlikeness, $r(106) = .24$, $p < .05$, in RT.

ID analyses. In order to have a single measure of performance on the ID analyses that accounted for variability in both speed and accuracy, inverse efficiency scores were used (Table 5). These scores were calculated by dividing the time to complete each task by accuracy. Performance varied widely in each of these measures (Figure 8). Generally, participants who performed well on one ID measure tended to do so on the others as well. Performance on the spelling test was related to all other measures including author recognition, $r(106) = .36$, $p < .01$, vocabulary, $r(106) = .38$, $p < .01$, pseudoword naming, $r(106) = .29$, $p < .01$, and sight word naming, $r(106) = .35$, $p < .01$. Performance on the vocabulary test was related to author recognition, $r(106) = .44$, $p < .01$ (Figure 9, left). There was also a positive relationship the two TOWRE measures, $r(106) = .35$, $p < .01$ (Figure 9, right). The top section of Table 6 shows each of these correlations.

Relationships between the experimental tasks and ID measures. Correlations were performed between the ID measures, the experimental tasks, and effects. These are presented in the bottom section of Table 6. In general, participants who showed poorer performance on the ID measures were also slower to respond and had larger effects in RT. Notably, however, there were fewer relationships between the ID measures and the semantic categorization task. Participants

who were slower to respond to the semantic categorization task performed somewhat more poorly on the spelling test, $r(106) = .21$, $p < .05$ (Figure 10, left). Those with larger homophone effects tended to perform more poorly on the author recognition task, $r(106) = .23$, $p < .01$. This is in contrast to the pattern seen with the lexical decision task. Here, participants who were slower to respond to the nonwords in the lexical decision task performed more poorly on the author recognition task, $r(106) = .35$, $p < .01$, spelling test, $r(106) = .49$, $p < .01$ (Figure 10, right), vocabulary test, $r(106) = .32$, $p < .01$, pseudoword naming task, $r(106) = .26$, $p < .01$, and sight word naming task, $r(106) = .40$, $p < .01$. A similar pattern was seen in response to the words from the lexical decision task, as participants who responded more slowly performed more poorly on the author recognition task, $r(106) = .26$, $p < .01$, spelling test, $r(106) = .42$, $p < .01$, vocabulary test, $r(106) = .27$, $p < .01$, and sight word naming task, $r(106) = .34$, $p < .01$, but not quite the pseudoword naming task, $r(106) = .15$, *ns*. Participants with larger wordlikeness effects in RT performed more poorly on all ID measures, including author recognition, $r(106) = .20$, $p < .05$, spelling, $r(106) = .25$, $p < .01$, vocabulary, $r(106) = .31$, $p < .01$, pseudoword naming, $r(106) = .18$, $p < .05$, and sight word naming, $r(106) = .33$, $p < .01$. Similarly, those with larger effects of pseudohomophony in RT also performed more poorly in spelling, $r(106) = .32$, $p < .01$, vocabulary, $r(106) = .30$, $p < .01$, pseudoword naming, $r(106) = .27$, $p < .01$, and sight word naming, $r(106) = .25$, $p < .01$, but not quite author recognition, $r(106) = .15$, *ns*. Likewise, participants with a larger effect of imageability in RT performed more poorly on the author recognition task, $r(106) = .22$, $p < .01$, spelling test, $r(106) = .31$, $p < .01$, vocabulary test, $r(106) = .32$, $p < .01$, and sight word naming task, $r(106) = .22$, $p < .01$, but not the pseudoword naming task, $r(106) = .08$, *ns*.

With few exceptions, the correlations between the measures in the ID battery and error rates from the experimental tasks are most notable for their absence. Only vocabulary and pseudoword naming showed relationships with any of the intercepts or effects from the models of error rate. Participants who performed poorly in pseudoword naming made somewhat more errors on the semantic categorization task, $r(106) = .22$, $p < .05$, and in response to the nonwords from the lexical decision task, $r(106) = .28$, $p < .01$. Those who performed poorly on the vocabulary test made more errors on the semantic categorization task, $r(106) = .21$, $p < .05$. Finally, there was only one significant negative correlation between an ID measure and an experimental effect. Participants who performed better on the vocabulary test also had larger effects of pseudohomophony, $r(106) = -.19$, $p < .05$.

Discussion

As predicted, all the psycholinguistic effects found replicate the group-level effects established in previous literature. The homophone effect found in the semantic categorization task replicates the findings of Van Orden (1987) and successive papers, with participants making more errors and responding more slowly to stimuli that are homophones of category exemplars. In accordance with the results of Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004), participants also responded more slowly to less imageable words. As in Coltheart, Davelaar, Jonasson, and Besner (1977), participants made more errors and responded more slowly to more wordlike nonwords in a lexical decision task. The effect of pseudohomophony was very similar to that of wordlikeness, with participants making more errors and responding more slowly to pseudohomophones than non-homophonic nonwords. Having established these main effects, I now consider the individual variability in each effect.

I suggested these effects and their interactions might provide insight into individual differences in division of labor. The mixed effects models used to characterize individual task success and effect sizes proved an excellent measure. While they also proved useful in measuring the experimental effects in RT, they were less successful in some of the error rates. This was especially true of the imageability effect from the model, as no direct relationship can be seen with the imageability effect from difference scores. The relationship between the wordlikeness effect from the model and from difference scores was also rather weak, although this may have been the result of including individual effects of pseudohomophony and wordlikeness in the same model. The wordlikeness effect also showed the smallest amount of variability in error rates, lending support to this possibility. Variability in the homophone effect in error rates was greater than that in the wordlikeness effect, but smaller than those in the imageability and pseudohomophone effects. In RT, the homophone effect showed the smallest variability, which may explain why relationships with this effect were not as strong as with the other effects.

It was suggested that participants with a large homophone or pseudohomophone effect, small imageability effect, or poor spelling may primarily use the phonological pathway, whereas those showing the opposite pattern would rely more on the semantic pathway. However, the homophone effect was not related to any of the other effects in RT. However, in error rates, participants who had larger effects of homophony also tended to have larger effects of imageability and pseudohomophony, but smaller effects of wordlikeness. There was a strong negative relationship between the pseudohomophone and wordlikeness effects as well. One possible reason for this could be that orthographic information is processed rapidly, so that those with large wordlikeness effects make more errors on both homophonic stimuli and controls, resulting in smaller effects of homophony and pseudohomophony. Alternatively, this may be a

glimpse of division of labor between the semantic and phonological pathways. I predicted that participants with larger effects of wordlikeness could rely more on the semantic pathway while those with larger effects of homophony and pseudohomophony could be biased toward the phonological pathway. While there is some evidence for this, an examination using other psycholinguistic effects would be necessary to confirm this theory.

I predicted that there would be systematic relationships among these experimental effects and with the measures from the ID battery. This proved true to some degree. Participants who performed better on the Author Recognition Task tended to have smaller homophone effects. As the Author Recognition Task is designed to index reading experience, it may be that the homophone effect shrinks as individuals gain reading experience. This question would benefit from a sample of participants with greater variability in reading experience. Interestingly, there were no relationships with either the spelling or vocabulary tests, suggesting that the homophone effect is not being driven by participants not knowing the correct spellings of nor being unfamiliar with the stimuli.

While relationships with the homophone effect were limited in this study, this was not true of all the experimental effects. With regards to the effects drawn from the lexical decision task, participants with one large effect often showed other large effects in RT. Relationships in the error rates were quite limited, perhaps due to their weaker relationships to the raw data. This pattern continued as I examined relationships with both experimental and ID task performance. Most of the effects in RT were found to be related to the ID measures, while this was not true in error rates. Participants with larger effects in RT often respond more slowly to each task, but this may be due to the wider range of responses or to floor effects in RT. While the data suggests participants with larger effects in error rates also made fewer errors on most tasks, these

relationships tended to be weaker. This interesting difference between how effects are related to RT and error rates may be due to participants who make few errors doing so only on difficult items, with participants who make more errors doing so on all types of items. Ultimately, the data suggests that individual differences may be easier to identify in RT than in error rates. Possible reasons for this include the skewed nature of the raw error rates, smaller variability in error rates, and the type of model used to quantify these effects.

It was also predicted that the homophone effect in semantic categorization would prove a useful marker of relative reliance on phonological pathway. As success on the semantic categorization task and the effect of homophony in RT do not seem related to the ID measures, it could be the case that the semantic categorization task is ill suited for picking up individual differences. Participants who responded faster on one task tended to be faster on the others, however, this relationship may be due to a general speed factor rather than something task related. In general, our data suggests that the degree to which individual differences are detected can depend upon both the nature of the task and the measures being collected.

Limitations of the Current Study

One area in which I was unable to cast a wide net was with our participants, all of whom were university undergraduates. The relative homogeneity of our sample may have limited the variability in experimental tasks and effect sizes. The inclusion of participants from other or additional populations could results in the identification of additional differences. Nevertheless, the suggestion of individual differences with a constrained sample may mean that there are even greater differences in the larger population. Nevertheless, determining that these individual differences are stable and inherent to participants may require a longer study with task repetition over time. Due to time constraints, this was beyond the scope of this experiment.

Our investigation of division of labor would have benefited from the inclusion of additional tasks, but this study was limited by the amount of time needed to conduct such an experiment. Including a naming task to capture individual differences in imageability may have provided a clearer comparison of the two pathways. In addition, a domain general measure of response time would have been beneficial in determining how much of the relationships in RTs may have been driven by such a factor. Another consideration is that performing lexical decision and semantic categorization tasks are unlikely to be identical to reading. However, similar patterns of brain activation have been found for reading aloud and lexical decision (Carreiras, Mechelli, Estevez, & Price, 2007), but if this proves to be a concern, it may be worthwhile to examine individual differences using tasks such as eye tracking. In any case, the individual differences literature will only benefit from the inclusion of additional types of measures.

Future Directions

This study examined the effects of homophony, wordlikeness, imageability, and pseudohomophony to better understand how people differ with regards to each of these effects and their relationship to success in semantic categorization and lexical decision tasks. Future studies should replicate these results using additional and different psycholinguistic effects, such as concreteness, priming, transposed letter, and neighborhood effects. As mentioned above, such replications should include additional types of tasks, such as naming and eye tracking, as well as studies designed to see the degree to which these individual differences are stable over time. These results should also be replicated in different populations and across languages, as the size and directionality of these effects can differ greatly based upon the nature of the participants and the structure of the language. Further, while the five measures in our ID battery were insightful, additional and more extensive measures, such as a reading comprehension and working memory,

would improve our ability to create a reading profile for each reader. Concurrent with this behavioral research, it would also be beneficial to extend the work of Harm and Seidenberg (2004) by developing computational models that account for all the systematic variation discovered with regards to individual differences.

References

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*(6), 565–575. https://doi.org/10.3758/BF03202439

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 386–393. https://doi.org/10.1037/0096-1523.2.3.386

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 537–554. https://doi.org/10.1037/0096-1523.20.3.537

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*(2), 68–76. https://doi.org/10.3758/BF03197587

Carreiras, M., Mechelli, A., Estevez, A., & Price, C. J. (2007). Brain activation for lexical decision and reading aloud: Two sides of the same coin? *Journal of Cognitive Neuroscience, 19*(3), 433-444. doi:10.1162/jocn.2007.19.3.433

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hilldale, NJ: Erlbaum.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches,. *Psychological Review*, *100*(4), 589–608. https://doi.org/10.1037/0033-295X.100.4.589

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing in children: A quick measure of reading experience. *Journal of Educational Psychology, 82*(4), 733.

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review, 19*, 325-331. doi:10.3758/s13423-011-0213-7

Grainger, J., Lete, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662-720. doi:10.1037/0033-295X.111.3.662

Herdman, C. M., & LeFevre, J. (1992). Individual differences in the efficiency of word recognition. *Journal of Educational Psychology, 84*(1), 95-102.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(2), 385–402. https://doi.org/10.1037/a0038995

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and toward logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi:10.1016/j.jml.2007.11.007

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128*(3), 219-264.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*(4), 358-394. doi:10.1037/0096-3445.120.4.358

Jorm, A. F. (1977). Effect of word imagery on reading performance as a function of reader ability. *Journal of Educational Psychology, 69*(1), 46-54.

Kuperman, V., Ls, C., Dyke, J. a Van, Laboratories, H., Street, G., & Haven, N. (2007). Individual differences in visual comprehension of morphological complexity, 1643–1648.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. https://CRAN.R-project.org/package=lmerTest

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122*(3), 316-330.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*, 203-208.

Nelson, M. J., Brown, J. I., & Denny, M. J. (1960). *The Nelson-Denny Reading Test: Vocabulary, Comprehension, Rate.* Houghton Mifflin.

Newman, R. L., & Joanisse, M. F. (2011). Modulation of brain regions involved in word recognition by homophonous stimuli: An fMRI study. *Brain Research*, *1367*, 250–264. https://doi.org/10.1016/j.brainres.2010.09.089

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 139-156. doi:10.1037/0278-7393.27.1.139

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–68. https://doi.org/10.1037/0033-295X.96.4.523

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition*, *6*(4), 410–415. https://doi.org/10.3758/BF03197473

Starr, M. S., & Fleming, K. K. (2001). A rose by any other name is not the same: The role of orthographic knowledge in homophone confusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 744-760.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*(4), 347–359. https://doi.org/10.1037/h0087322

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1140-1154.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency-Second Edition (TOWRE-2).* Austin, TX: Pro-Ed.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181-198. doi:10.3758/BF03197716

Weber, R. M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly,* 5, 427-451.

Woollams, A. M., Lambon Ralph, M. A., Madrid, G., & Patterson, K. E. (2016). Do You Read How I Read? Systematic Individual Differences in Semantic Reliance amongst Normal Readers. *Frontiers in Psychology*, *7*(November), 1–16. https://doi.org/10.3389/fpsyg.2016.01757

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single-versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology-Human Perception and Performance*, *32*(6), 1324–1344. https://doi.org/10.1037/0096-1523.32.6.1324

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology-Human Perception and Performance*, *38*(1), 53–79. https://doi.org/10.1037/a0024177

Tables

Table 1

Means RT and error rates

Task	Measure	Conditions				Effect
SCT						
		Homophones	(SD)	Spelling Controls	(SD)	Homophony
	Mean RT	1257	(246)	1166	(243)	91
	Mean ER	.285	(.083)	.130	(.083)	.155
LDT						
		Pseudo-homophonic nonwords	(SD)	Non-homophonic nonwords	(SD)	Pseudohomophony
	Mean RT	765	(132)	727	(119)	38
	Mean ER	.102	(.096)	.040	(.054)	.062
		More wordlike nonwords		Less wordlike nonwords	(SD)	Wordlikeness
	Mean RT	776	(133)	717	(116)	59
	Mean ER	.099	(.090)	.043	(.055)	.056
		Low imageability words		High imageability words	(SD)	Imageability
	Mean RT	709	(89)	658	(77)	51
	Mean ER	.166	(.076)	.038	(.047)	.128

Note: Mean RT (ms) and error rate are the means of participant means.

Table 2

Model comparison

Model	b	X^2
Semantic Categorization		
RT ~ controls + homophony	0.073	6.9**
RT ~ controls + homophony + (1+homophony|subject)		895.8***
ER ~ controls + homophony	1.088	10.9***
ER ~ controls + homophony + (1+homophony|subject)		61.6***
Lexical Decision (words)		
RT ~ controls + imageability	0.089	18.3***
RT ~ controls + imageability + (1+imageability|subject)		774.4***
ER ~ controls + imageability	1.417	17.7***
ER ~ controls + imageability + (1+imageability|subject)		120.4***
Lexical Decision (nonwords)		
RT ~ controls + pseudohomophony	0.052	11.6***
RT ~ controls + pseudohomophony + (1+pseudohomophony|subject)		2553.0***
RT ~ controls + wordlikeness	0.079	28.3***
RT ~ controls + wordlikeness + (1+wordlikeness|subject)		2542.0***
RT ~ controls + pseudohomophony…	0.052	43.2***
+ wordlikeness	0.079	
RT ~ controls + pseudohomophony + wordlikeness…		2564.5***
+ (1+pseudohomophony+wordlikeness|subject)		
ER ~ controls + pseudohomophony	0.890	18.7***
ER ~ controls + pseudohomophony + (1+pseudohomophony|subject)		510.7***
ER ~ controls + wordlikeness	0.897	19.0***
ER ~ controls + wordlikeness + (1+wordlikeness|subject)		497.0***
ER ~ controls + pseudohomophony…	0.896	41.1***
+ wordlikeness	0.894	
ER ~ controls + pseudohomophony + wordlikeness…		510.9***
+ (1+pseudohomophony+wordlikeness|subject)		

Note: * p < .05 ** p < .01 *** p < .001

Note: All models controlled for item, item order, and previous item RT.

Table 3

Correlations between means and models

Model	**r**
RT	
Semantic categorization	.998
Lexical decision task (words)	.997
Lexical decision task (nonwords)	.994
Error rates	
Semantic categorization	.995
Lexical decision task (words)	.929
Lexical decision task (nonwords)	.965
Effects in RT	
Homophone effect	.931
Pseudohomophone effect	.860
Wordlikeness effect	.808
Imageability effect	.754
Effects in error rates	
Homophone effect	.835
Pseudohomophone effect	.799
Wordlikeness effect	.360
Imageability effect	.084

Table 4

Correlations between and among performance and effect size

			Performance						**Effect Size**							
			RT			ER			RT				ER			
			SCT	LDT word	LDT NW	SCT	LDT word	LDT NW	HPH	IMG	PHP	WDL	HPH	IMG	PHP	WDL
Tasks	RT	SCT	1.00													
		LDT word	**0.45**	1.00												
		LDT nonword	**0.53**	**0.75**	1.00											
	ER	SCT	**0.26**	0.14	**0.31**	1.00										
		LDT word	**-0.22**	0.01	**-0.22**	0.06	1.00									
		LDT nonword	0.06	-0.02	**0.24**	**0.33**	**0.22**	1.00								
Effects	RT	Homophony	**0.30**	0.14	0.08	-0.09	**-0.25**	**-0.20**	1.00							
		Imageability	**0.34**	**0.75**	**0.58**	0.09	-0.07	-0.05	0.10	1.00						
		Pseudohomophony	**0.31**	**0.54**	**0.59**	**0.21**	**-0.17**	0.06	0.08	**0.47**	1.00					
		Wordlikeness	**0.31**	**0.45**	**0.59**	**0.17**	**-0.24**	0.10	0.10	**0.48**	**0.78**	1.00				
	ER	Homophony	0.05	**0.17**	0.11	**-0.29**	-0.12	-0.07	**0.23**	0.13	**0.26**	**0.25**	1.00			
		Imageability	**0.21**	0.07	**0.21**	-0.07	**-0.91**	**-0.24**	**0.24**	**0.16**	**0.18**	**0.24**	**0.16**	1.00		
		Pseudohomophony	0.02	-0.04	-0.05	0.00	-0.11	0.03	0.03	-0.03	0.10	0.09	**0.24**	0.12	1.00	
		Wordlikeness	-0.04	-0.01	-0.07	-0.06	0.03	**-0.25**	0.08	0.03	-0.01	-0.06	**-0.21**	-0.08	**-0.79**	1.00

Note: Correlations are between the individual participant random intercepts from each of the mixed effects models. RT = response time; ER = error rate; NW = nonword; HPH = homophony; IMG = imageability; PHP = pseudohomophony; WDL = wordlikeness.

Table 5

Mean inverse efficiency scores for ID measures

Task	**Mean**	**(SD)**
Author Recognition Task	3.95	(1.19)
Nelson-Denny Vocabulary Test	11.40	(4.36)
Spelling Test	4.17	(1.26)
TOWRE Pseudoword Reading	0.91	(0.20)
TOWRE Sight Word Reading	0.49	(0.07)

Note: Means and SDs are of the individual inverse efficiency scores, calculated by dividing the task completion time by accuracy.

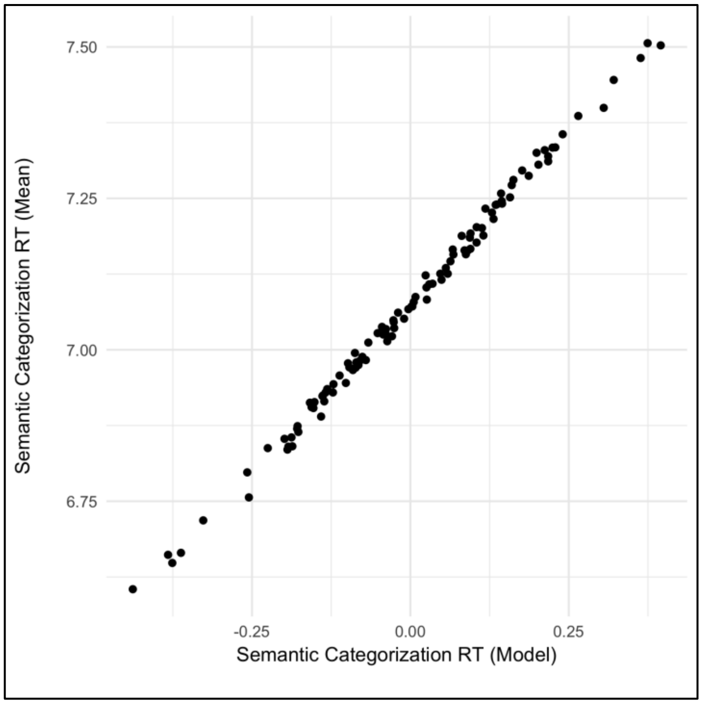
Table 6

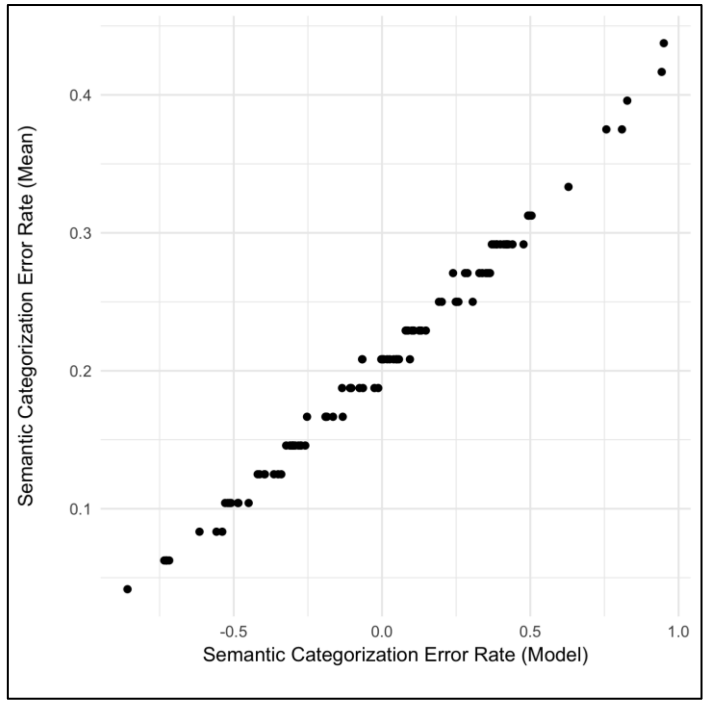
Correlations between the ID and experimental tasks

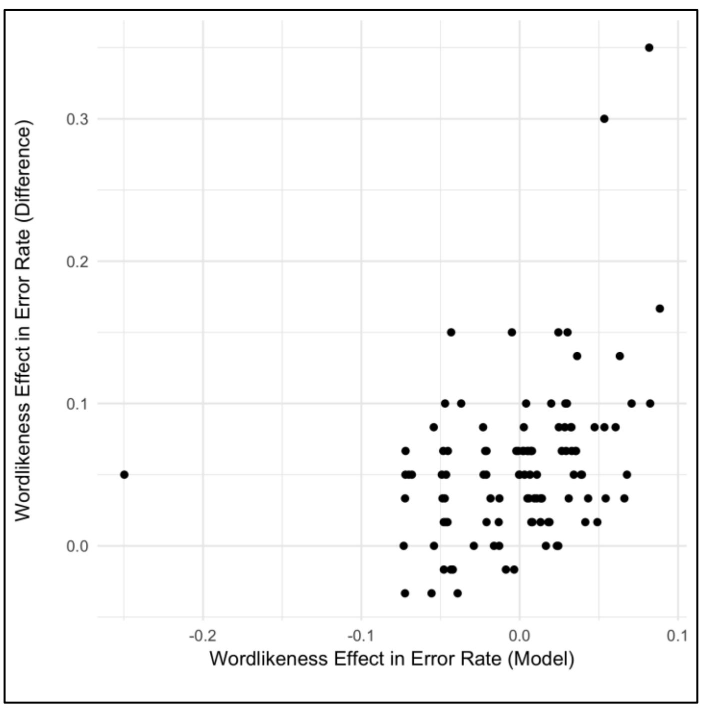
	Measure	ART	SPL	VOC	PDE	SWE
	Author Recognition Task	1.00				
	Spelling Test	**0.44**	1.00			
	Nelson-Denny Vocabulary Test	**0.36**	**0.38**	1.00		
	TOWRE Pseudoword Naming	0.10	**0.29**	**0.16**	1.00	
	TOWRE Sight word Naming	0.15	**0.35**	0.08	**0.35**	1.00
		ART	SPL	VOC	PDE	SWE
RT	SCT	0.14	**0.21**	0.14	-0.09	0.15
	LDT words	**0.26**	**0.42**	**0.27**	0.15	**0.34**
	LDT nonwords	**0.35**	**0.49**	**0.32**	**0.26**	**0.40**
	Homophone effect	**0.23**	0.08	0.04	-0.15	0.03
	Imageability effect	**0.22**	**0.31**	**0.32**	0.08	**0.22**
	Pseudohomophone effect	0.15	**0.32**	**0.30**	**0.27**	**0.25**
	Wordlikeness effect	**0.20**	**0.25**	**0.31**	**0.18**	**0.33**
ER	SCT	0.05	0.13	**0.21**	**0.22**	0.08
	LDT words	-0.12	0.02	-0.01	0.05	0.09
	LDT nonwords	-0.01	-0.01	0.14	**0.28**	0.14
	Homophone effect	0.01	0.09	0.02	0.02	0.14
	Imageability effect	0.14	-0.07	0.04	-0.08	-0.07
	Pseudohomophone effect	-0.11	-0.15	**-0.19**	-0.06	0.04
	Wordlikeness effect	0.12	0.11	0.14	0.05	-0.08

Note: Correlations are between the individual participant inverse efficiency scores on the ID battery and individual success and effect sizes from the experimental tasks. ART = Author Recognition Task; SPL = spelling test; VOC = Nelson-Denny Vocabulary Test; PDE = TOWRE Pseudoword Reading; SWE = TOWRE Sight Word Reading.

Figures


Semantic Categorization RT (Mean)
7.50
7.25
7.00
6.75
-0.25
0.00
0.25
Semantic Categorization RT (Model)


Semantic Categorization Error Rate (Mean)
0.4
0.3
0.2
0.1
-0.5
0.0
0.5
1.0
Semantic Categorization Error Rate (Model)


Wordlikeness Effect in Error Rate (Difference)
0.3
0.2
0.1
0.0
-0.2
-0.1
0.0
0.1
Wordlikeness Effect in Error Rate (Model)

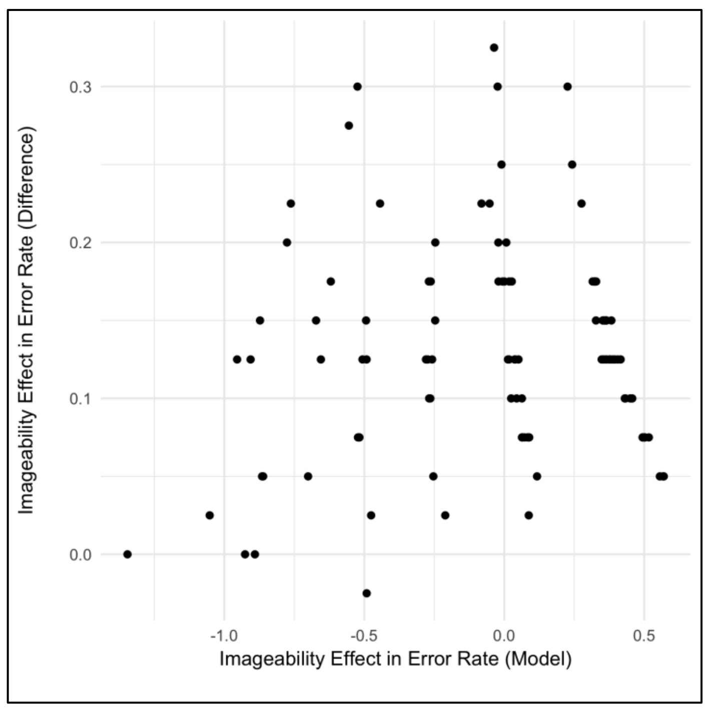

Imageability Effect in Error Rate (Difference)
0.3
0.2
0.1
0.0
-1.0
-0.5
0.0
0.5
Imageability Effect in Error Rate (Model)

Figure 1. Clockwise from top-left: Relationships in model predictions and mean of means for SCT (1) RT (r^2=0.998) and (2) ER (r^2=0.931), and difference scores for (3) imageability (r^2=0.084) and wordlikeness (r^2=0.360) effects in ER.

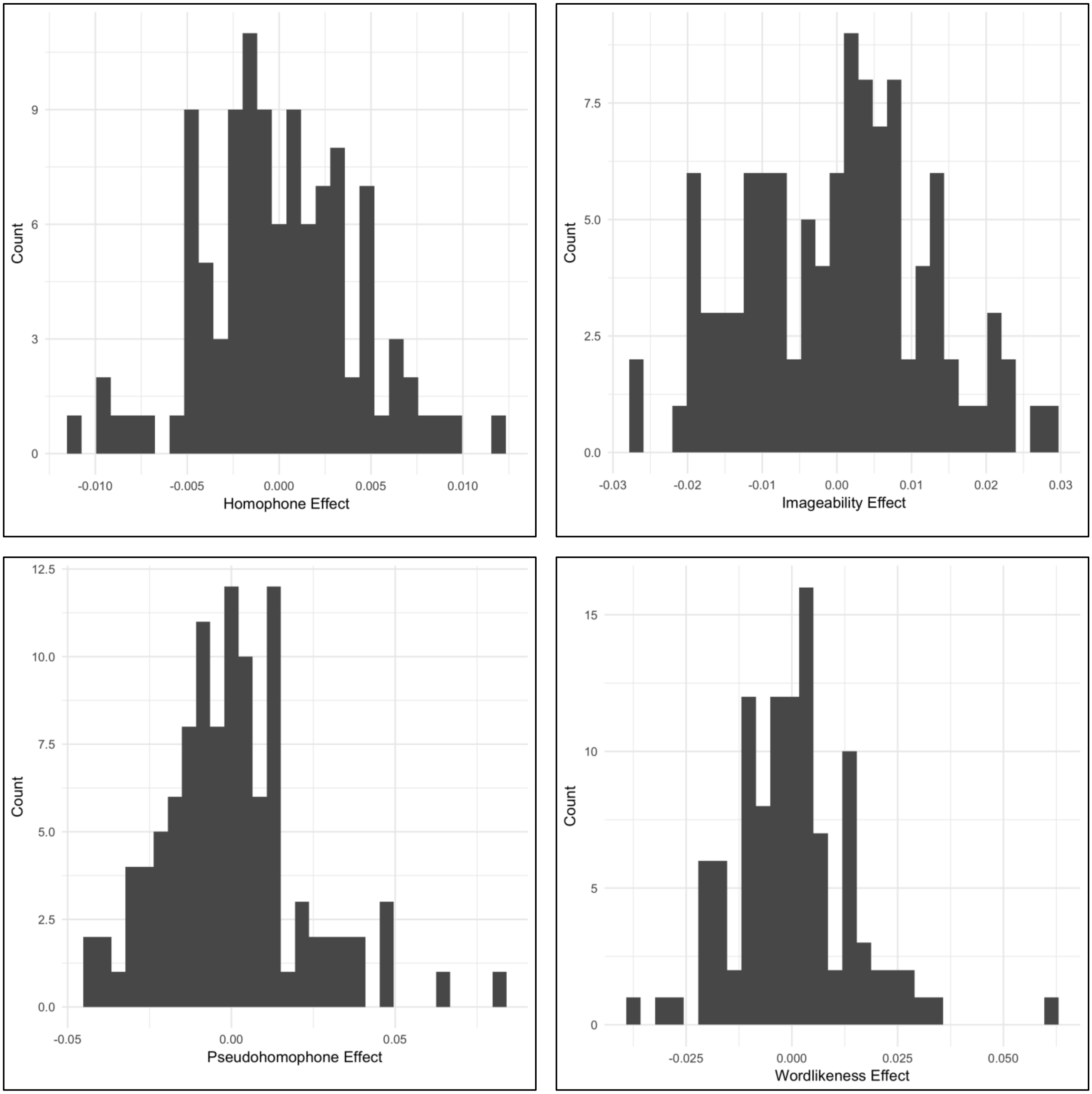


Figure 2. Individual variability in the experimental effects (RT).

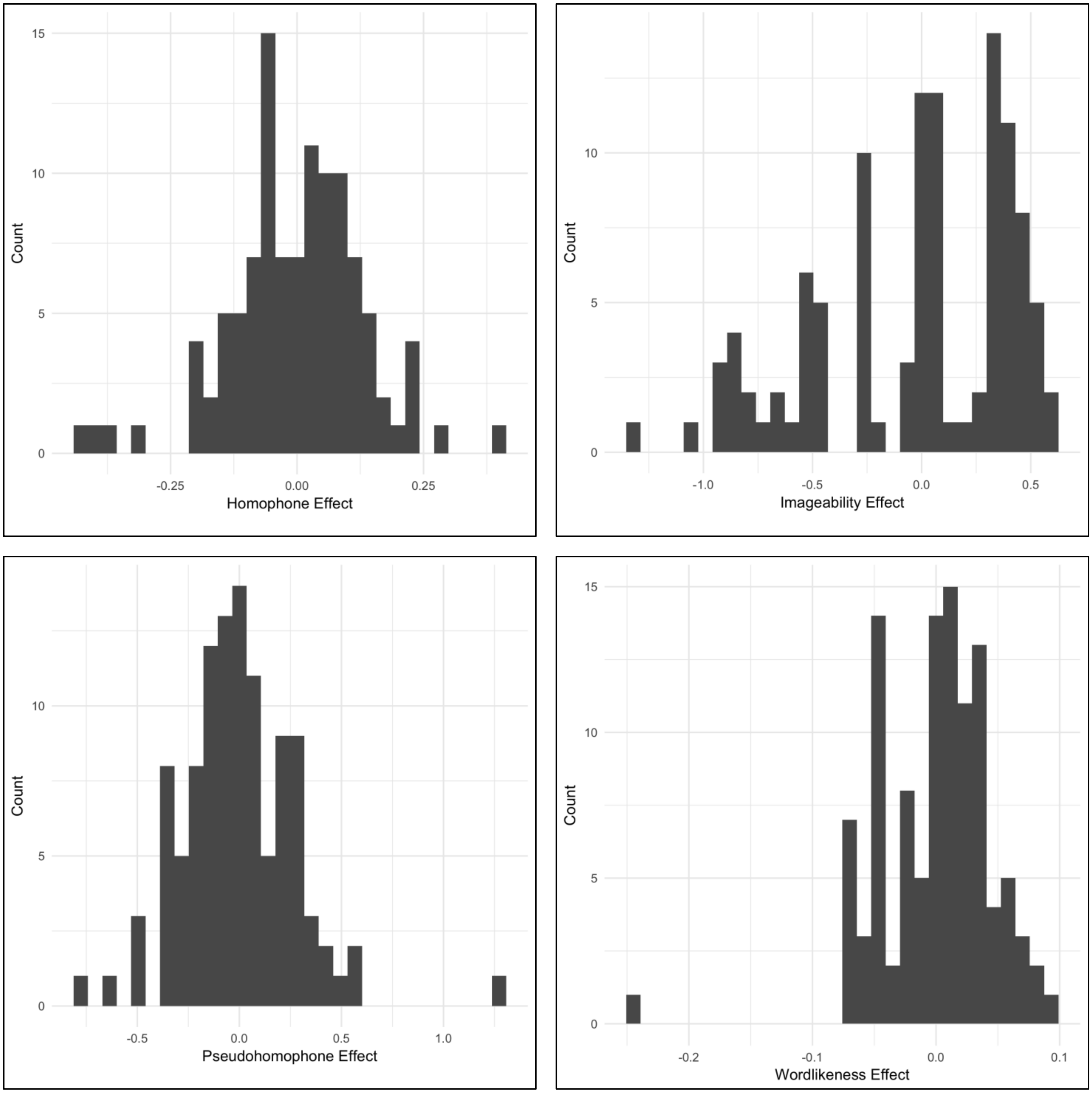


Figure 3. Individual variability in the experimental effects (error rate).

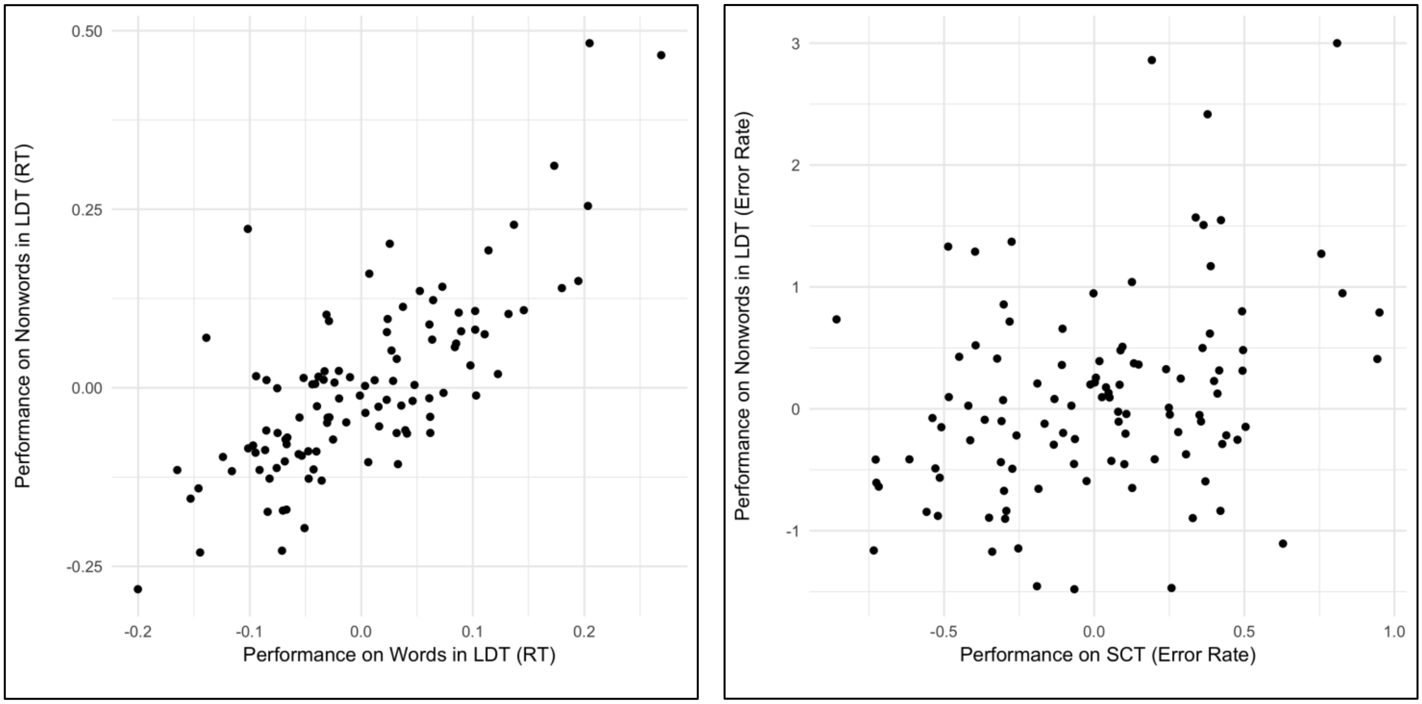
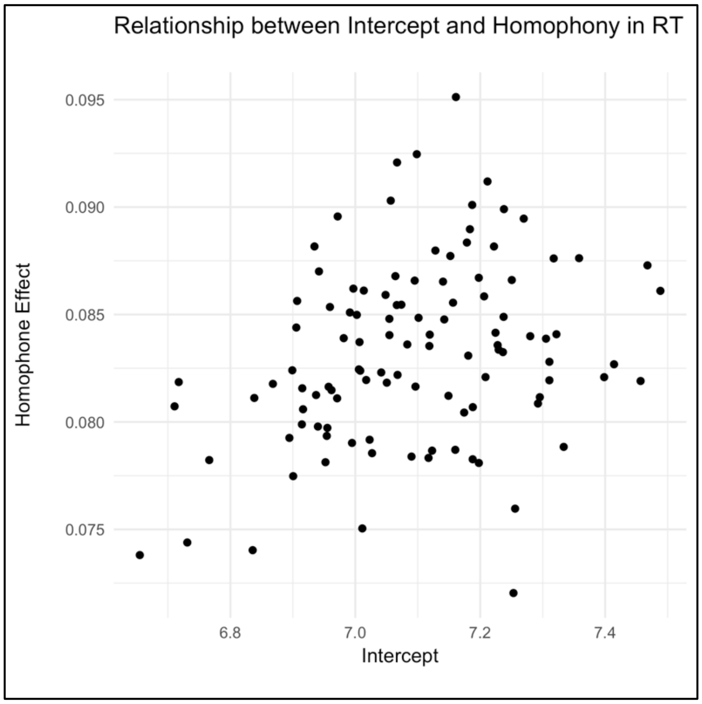
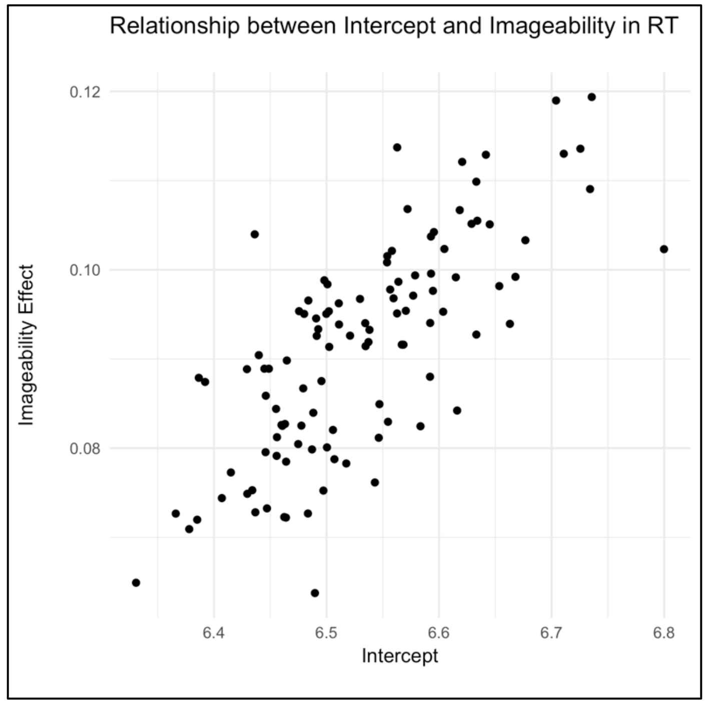
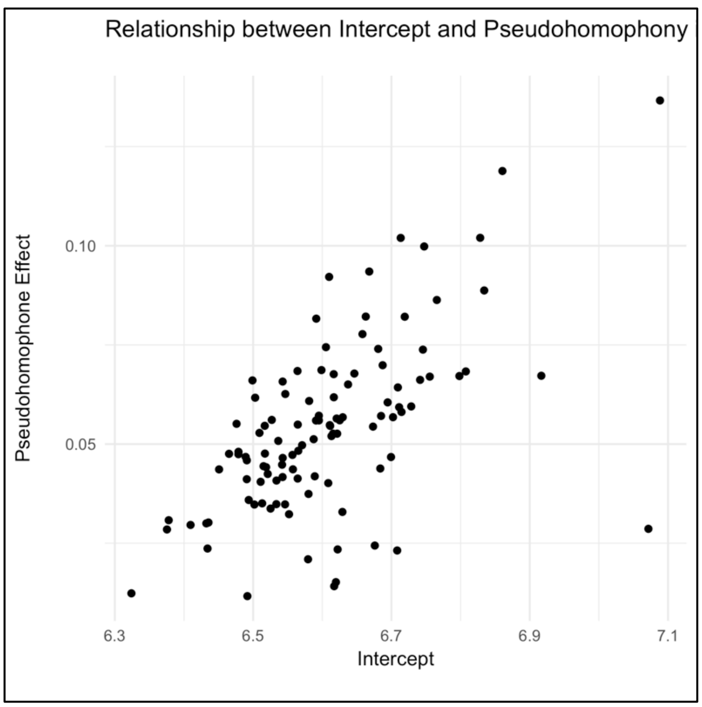


Figure 4. Left: Relationship between LDT words and nonwords in RT (r^2=0.75); right: Relationship between SCT and LDT nonwords in ER (r^2=0.22).


Relationship between Intercept and Homophony in RT
Homophone Effect
0.095
0.090
0.085
0.080
0.075
6.8
7.0
7.2
7.4
Intercept


Relationship between Intercept and Imageability in RT
Imageability Effect
0.12
0.10
0.08
6.4
6.5
6.6
6.7
6.8
Intercept


Relationship between Intercept and Pseudohomophony
Pseudohomophone Effect
0.10
0.05
6.3
6.5
6.7
6.9
7.1
Intercept

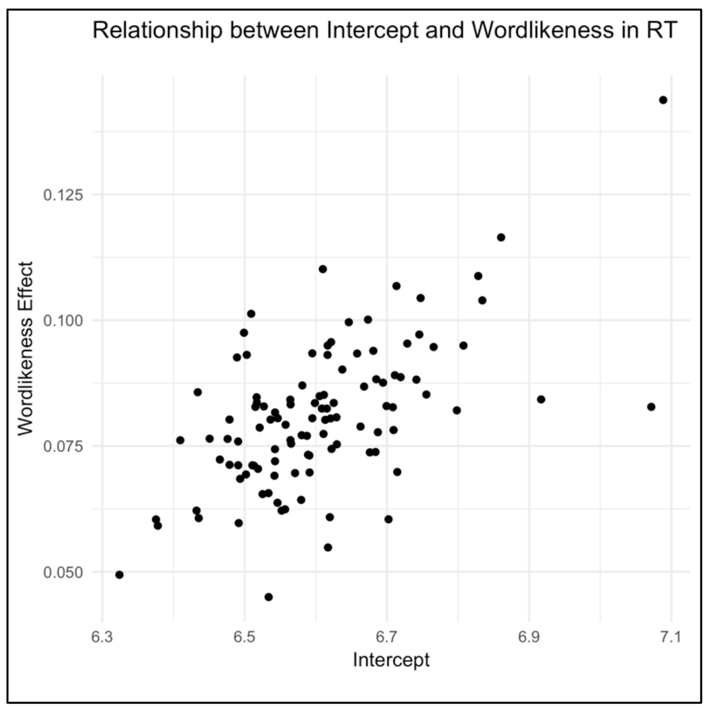

Relationship between Intercept and Wordlikeness in RT
Wordlikeness Effect
0.125
0.100
0.075
0.050
6.3
6.5
6.7
6.9
7.1
Intercept

Figure 5. Clockwise from top-left: Relationships in RT between task intercepts and the (1) homophone (r^2=0.30), (2) imageability (r^2=0.75), (3) wordlikeness (r^2=0.59), and pseudohomophone (r^2=0.59) effects.

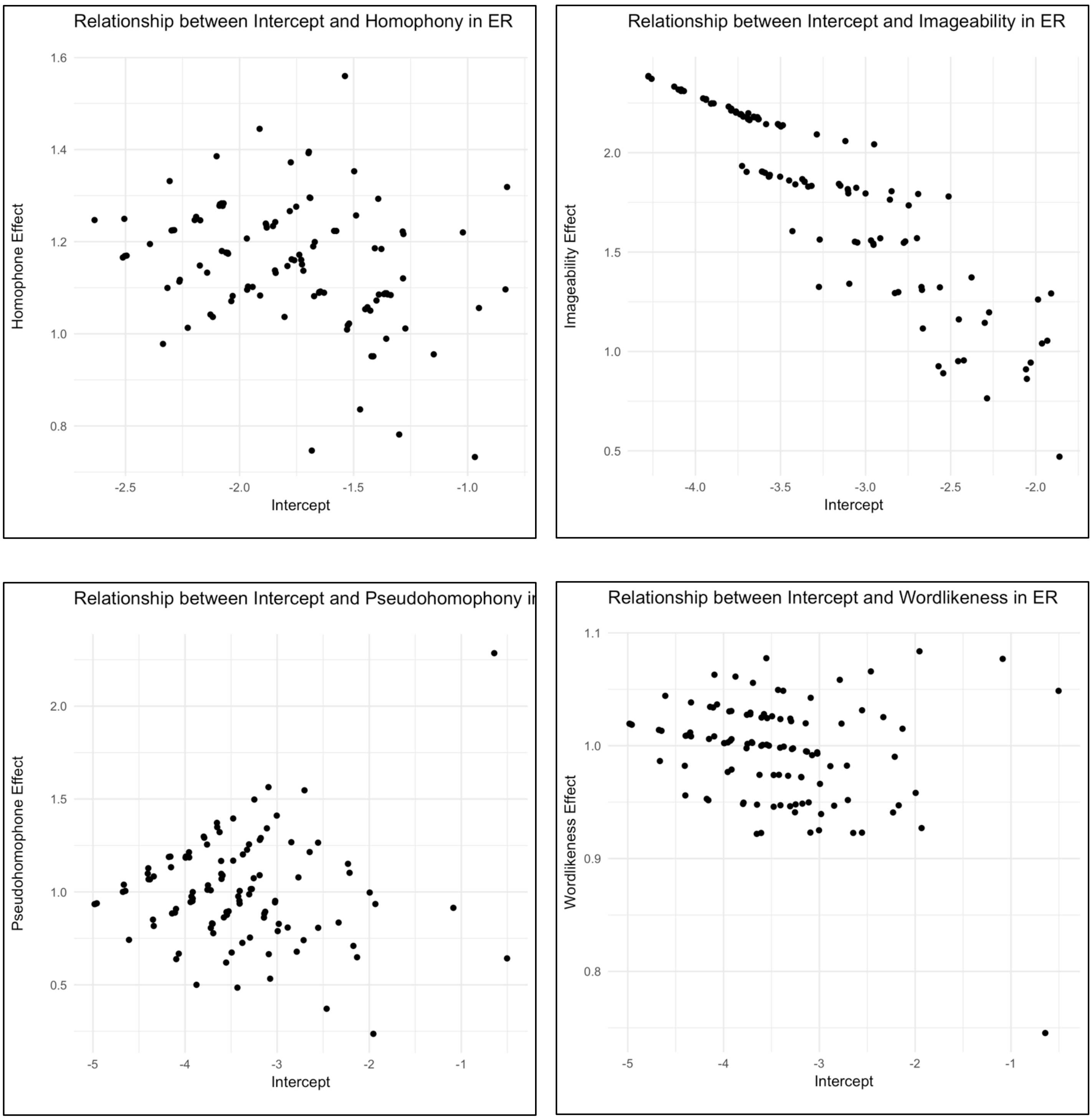


Figure 6. Clockwise from top-left: Relationships in ER between task intercepts and the (1) homophone (r^2=-0.29), (2) imageability (r^2=-0.91), (3) wordlikeness (r^2=-0.25), and pseudohomophone (r^2=0.21) effects.

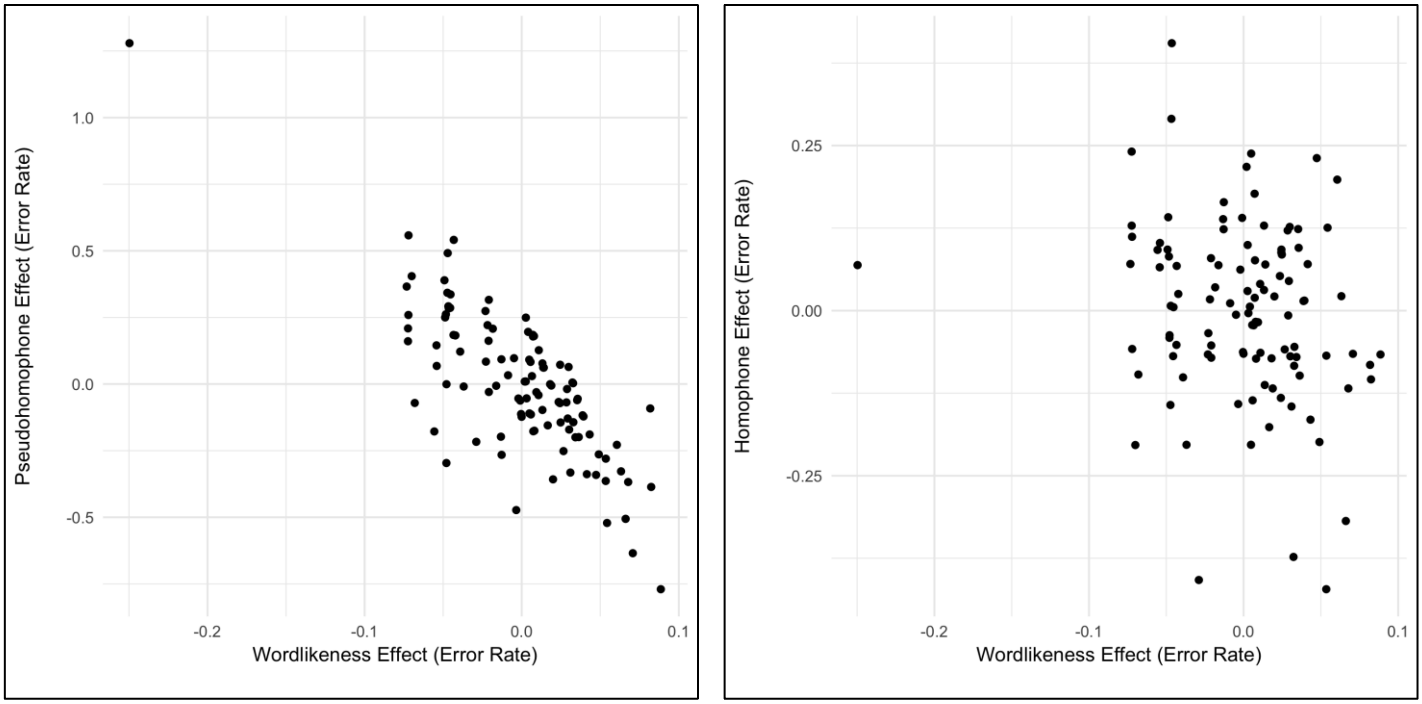


Figure 7. Left: Relationship the wordlikeness and pseudohomophone effects (r^2=-0.79) in error rates; right: Relationship between wordlikeness and homophone effects (r^2=-0.21) in error rates.

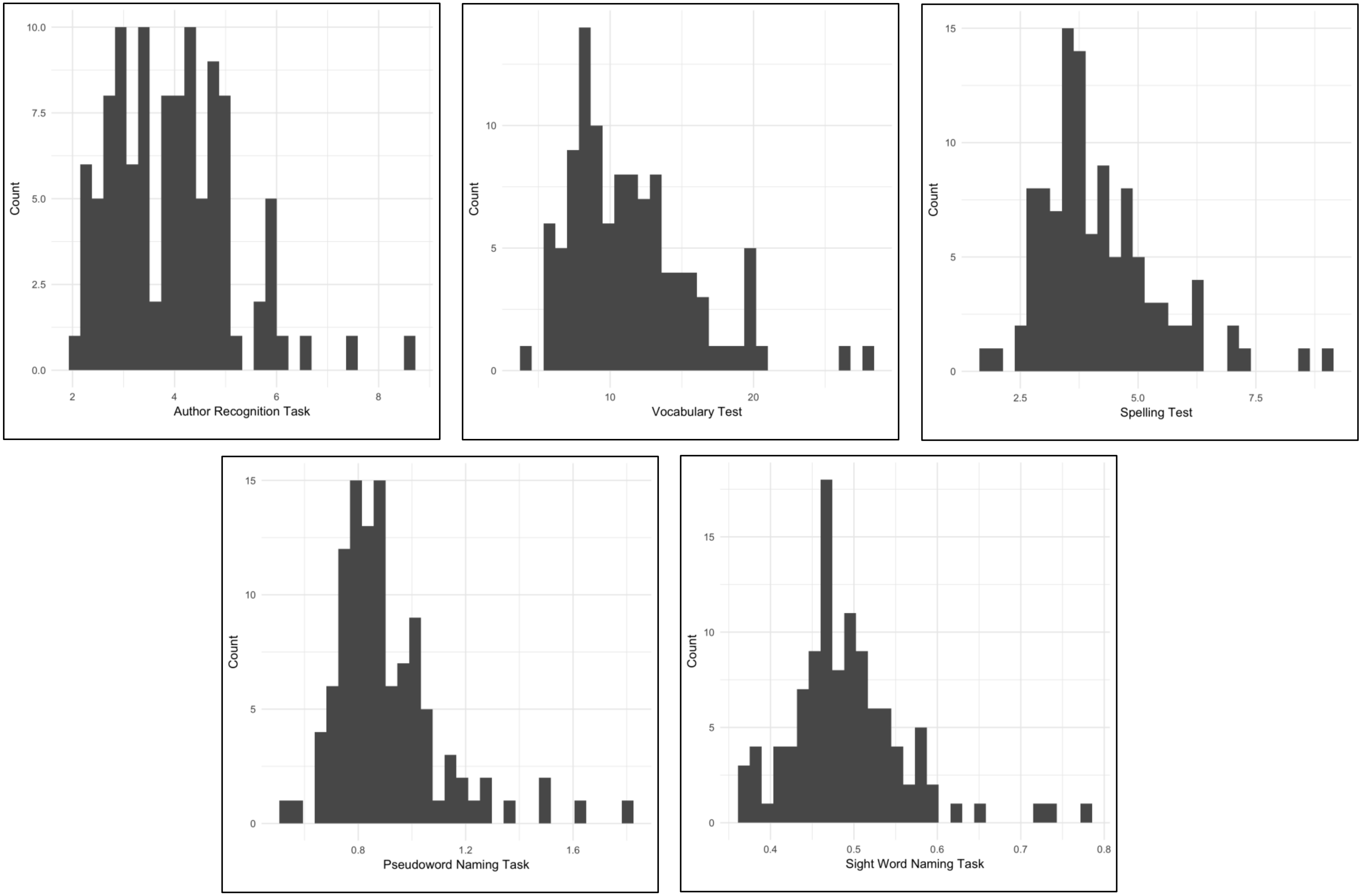


Figure 8. Variability in the ID measures.

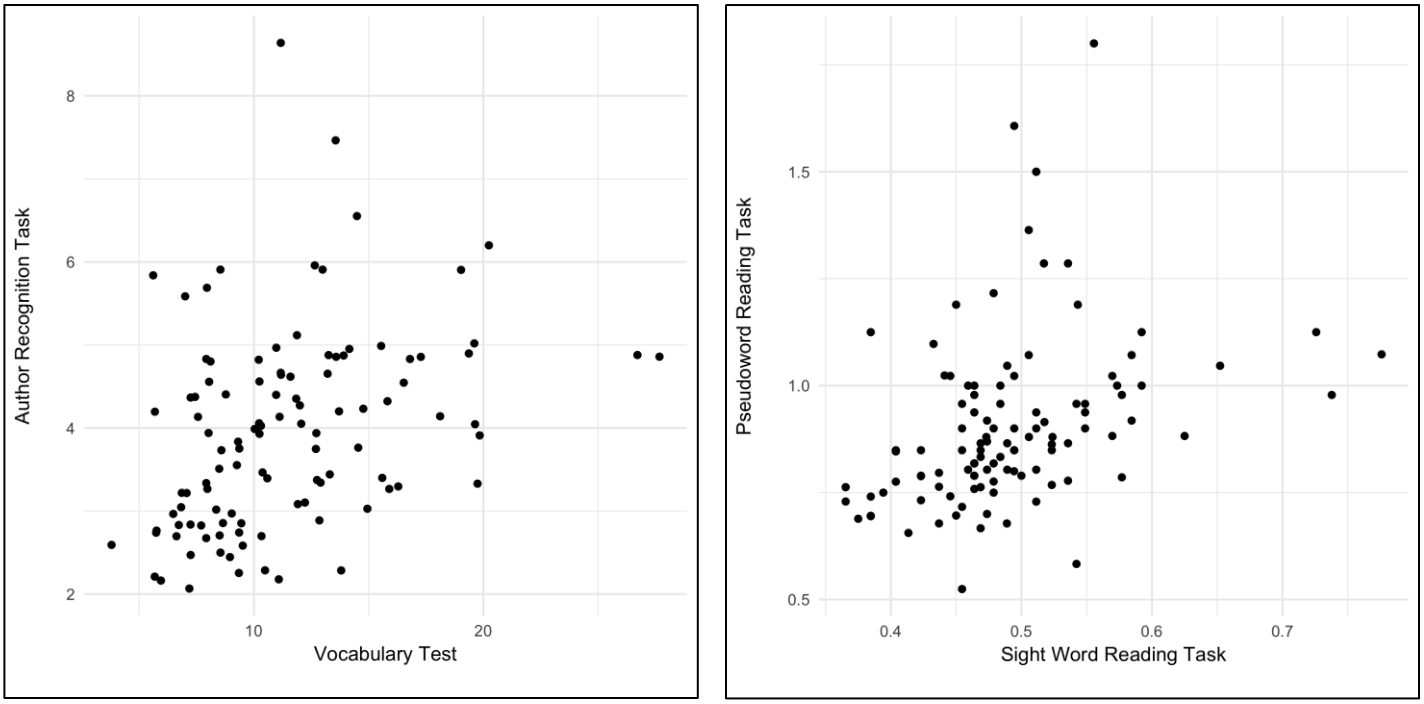


Figure 9. Left: Relationship between ART and vocabulary (r^2=0.36); right: Relationship between TOWRE-2 subtests (r^2=0.35).

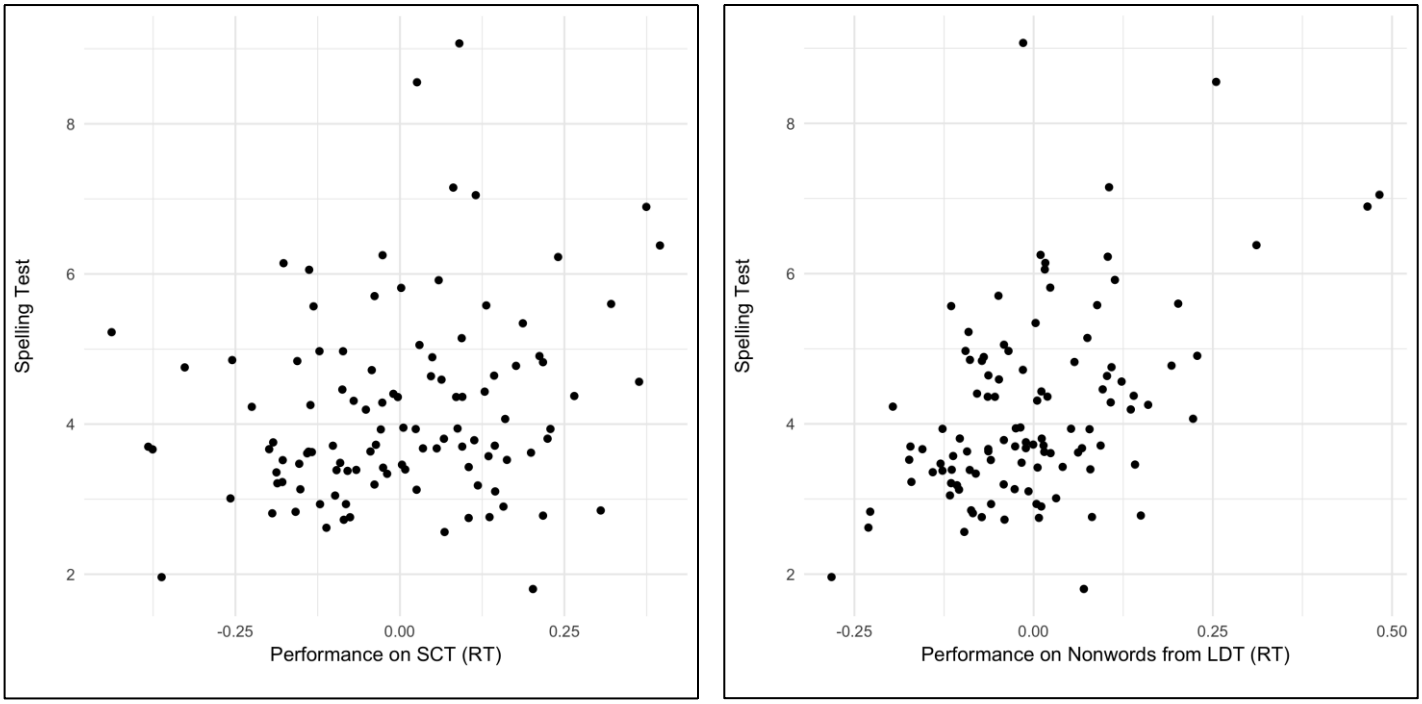


Figure 10. Left: Relationship between spelling and SCT (r^2=0.21); right: Relationship between spelling and LDT nonwords (r^2=0.49).