



Published in final edited form as:

Biotechniques. 1997 May ; 22(5): 952–957.

Ease of Articulation: A Replication

Linda I. Shuster, Ph.D., CCC/SLP¹ and Claire Cottrill

805 Allen Hall, West Virginia University, Morgantown, WV 26506, Phone: 304-293-4241

Linda I. Shuster: shusterl@gvsu.edu; Claire Cottrill: ccottri7@mix.wvu.edu

1. Introduction

In 2012, Begley and Ellis reported that Amgen, a major American biopharmaceutical company, had attempted to replicate the findings of 53 published cancer research studies that it deemed “high profile.” Amgen was able to reproduce the findings from only 6 of the 53 studies (11%). Concerns regarding reproducibility have been discussed among scientists for years, and increasingly more often in the popular press; however, the Amgen announcement led to increased interest and the creation of the Reproducibility Initiative, an effort to encourage authors of high profile research papers to allow independent investigators to replicate their findings. Recently, the U.S. National Institutes of Health (NIH) described initiatives that they will undertake to enhance reproducibility (Collins & Tabak, 2014). Articles regarding the problem of reproducibility have appeared in journals such as *Nature*, and, in 2012, the journal *Perspectives on Psychological Science* devoted an entire issue to the topic of reproducibility in psychology research. In the Introduction to that issue, the editors provide an excellent overview of the problems, including an unwillingness or inability to share published data, fewer replications than in the past, and questionable research practices (Pashler & Wagenmakers, 2012). However, some investigators have argued that the significance of the reproducibility problem has been exaggerated. Their main arguments to support this position are that 1) investigators use statistical methods to control the rate of false positives, 2) conceptual replications are conducted frequently, and 3) science is self-correcting (Pashler & Harris, 2012). Unfortunately, upon further scrutiny, these arguments do not hold up.

The first argument (statistical controls) does not hold up because the .05 probability level for statistical significance that is typically used in investigations in the social science disciplines does not represent the number of false positives throughout the discipline’s literature. In order to determine the number of literature-wide false positives, one must specify a post-

© 2015 Published by Elsevier Inc.

Correspondence to: Linda I. Shuster, shusterl@gvsu.edu.

¹Present address: 515 Michigan St. NE, Suite 200, Grand Valley State University, Grand Rapids, MI 49506, Phone: 616-331-5565

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interest: The authors report no conflict of interest. The authors alone are responsible for the content and writing of this paper.

study probability that an obtained effect is true, called the positive predictive value (PPV; Ioannidis, 2005). Based on the work of Ioannidis and other investigators and using a PPV of 10%, Pashler and Harris (2012) estimated that with an alpha of 5% and a power level of 80%, approximately 36% of published studies in psychology would be false positives. In that same journal, however, Bakker, van Dijk, and Wicherts (2012) estimated that the power level that is more typical of psychological studies is .35. Using the same procedure with a power level of .35, Pashler and Harris estimated a false positive rate that is considerably higher, approximately 56%.

The second argument (many conceptual replications) also does not hold up to scrutiny. A study of replications in psychology by Makel, Plucker, and Hegarty (2012) revealed that reproducibility was affected by the nature of the replication. If the replication was a direct replication by the original investigator(s) or a conceptual replication, the vast majority of studies reported findings that were similar to those of the original studies. A conceptual replication is one in which the rigor of the hypothesis is tested by employing different experimental methods. If, however, a direct replication (based on the methods reported in the original paper) was conducted by investigators who had no overlap with the original investigators, the study was significantly less likely to be successful in replicating the results of the original study.

The third argument (self correction) is also problematic. The argument in favor of self-correction is that because science is performed by making empirical observations, these observations can be confirmed or refuted by subsequent investigations. Those that are confirmed will stand, and those that are refuted will disappear from the literature. However, Ioannidis (2012) identified several impediments to self-correction in psychological and other sciences. These include publication bias (e.g., difficulties in getting negative results published), underpowered studies, lack of direct replications by independent investigators (due to bias toward direct replication inherent in the review process), and selective reporting bias. He argued that unless these biases are recognized and addressed, self-correction may not happen.

As noted above, in examining the top 100 journals in psychology, Makel et al. (2012) found few published reports involving direct replications. Based on a search of the literature, it seems that the same is true for journals related to communication sciences and disorders (CSD; Muma, 1993). Muma performed an analysis of the number of replications that had been published in *the Journal of Speech and Hearing Disorders* and the *Journal of Speech and Hearing Research* during the decade from 1979 to 1989. The combined data for both journals revealed only 9 direct replications out of a total of 271 studies. Based on these data, he estimated approximately 108 or 544 false findings, depending on the confidence interval employed. He also suggested that the rates of false positives might differ across the different communication disorders ("subpopulations"). Muma summarized the paper by arguing that: "there is an urgent need for more replications in the field of speech-language pathology and audiology (p. 929)."

The purpose of this paper is twofold. One purpose is to draw attention to the need for replication within the field of communication disorders. The second is to describe the

replication of a study that was reported in one of the journals evaluated by Muma (1993), the *Journal of Speech and Hearing Research*. In particular, we felt that a study performed by Locke (1972) regarding the relationship between ease of articulation and order of speech sound acquisition by children was worthy of replication. One reason was because of his experimental approach. Speech production is the most elegant and complicated motor behavior that humans produce (Kent, 2004). To explore motoric factors in speech sound acquisition, Locke examined the ease of articulation of American English speech sounds by exploiting the insight of mature speakers of the language to address the question of why children generally seem to acquire speech sounds in a particular order. He asked adults to rate how easy or difficult 20 consonants sounds were to produce. We found this to be an intriguing approach for assessing speech sound production difficulty. There are multiple technologies for studying the physiology of speech production, and, as Kent (1997) noted, they “permit the examination of even the most hidden aspects of speech (p. 306).” But there are no technologies that can observe the complex contribution of all of the speech subsystems simultaneously. As speakers, we experience the entire process of normal speech production, including respiratory, phonatory, resonatory, and articulatory aspects, as well as the demands for coordinating all of these aspects and the perceptual consequences of the movements. However, speech is a behavior that is acquired in an implicit, unconscious manner, so the validity of a task in which we try to gain conscious insight into this process is uncertain.

Another reason for replicating the Locke study was that it was designed to inform models and theories of speech sound acquisition in children. Locke found that the participants in his study rated certain sounds as being more difficult to produce than others. Moreover, he found that ease of articulation had a strong positive correlation with order of speech sound acquisition in children. Sounds that adults judged to be easier to produce were acquired earlier by children, and those judged to be more difficult were acquired later. In addition, he found a low correlation between children’s speech sound acquisition and their ability to recognize phonemes, thus he suggested that motoric factors are more important than perceptual factors in driving order of acquisition in American English. If Locke’s results can be replicated, particularly the ease ratings, it suggests that his tasks have some validity for assessing the relative difficulty of speech sound production in American English. It also suggests that the insight of mature speakers of the language might be useful for obtaining information about other aspects of communication behavior. Finally, it provides data that can be used to help inform models and theories of speech sound acquisition in children, and, by extension, models and theories of disordered speech. Locke reported the results of two studies in his article, and we replicated both of them.

Replication of Locke (1972) Experiment 1

2. Material and methods

2.1 Participants—Fifty-two undergraduate students participated in Experiment 1. They were all at least 18 years of age, were native speakers of English and had no history of previous speech or language problems.

2.2 Materials—Twenty word-initial phonemes were examined (η , ζ , θ , δ , were excluded; Table 1). Each phoneme was followed by “uh” to form a syllable and each syllable was paired with a different syllable (e.g., luh suh) in a written list containing 38 pairs. There were 5 different lists. The pairs were listed in random order, and no phoneme could succeed itself in the following pair (e.g., “ruh kuh, kuh vuh” was not allowed.). The creation of the materials followed Locke’s procedure.

2.3 Procedure—The lists were handed out in a junior-level speech-language pathology course. We used the same directions as in the Locke Experiment 1:

“Your job is to whisper each pair of sounds and try to decide which member of each pair is harder to say. “Harder” means it seems to require slightly more muscular effort or tension in the lips, jaw, tongue, or throat than the other member of the pair. Think only about the muscular activity in the lips, jaw, tongue, or throat. If you decide that an item is harder to say than the other, draw a circle around it. A sound may be harder in one case and easier in another, or it may always seem easier or harder, don’t look for patterns. Treat each pair as a separate case (p. 195).”

In addition, participants were told that if they were non-native speakers of English, they should put a ‘1’ at the top of their paper, and if they had a history of speech and/or language problems or therapy, they should put a ‘2’ at the top of the paper. This method for ensuring that participants had normal speech and were native speakers was not employed in the original study. Using the same procedure described in the original study, we tallied the number of times each phoneme (syllable) was encircled. We then used a Spearman rank order correlation to compare our data with Locke’s.

3. Results

Six participants were excluded who indicated that they had a history of previous speech therapy, leaving 52 participants. Columns 2 and 3 in table 1 show Locke’s data for the paired comparisons and the corresponding rank assigned to each phoneme based on the participants’ ratings, while columns 4 and 5 show our data for the paired comparisons and the corresponding phoneme rankings. The statistical analysis revealed a moderate (interpretation of strength based on Hinkle, Wiersma, & Jurs, 1979), significant positive correlation between our ranks and Locke’s ranks: $r_s=.69$, $t(18)=-4.05$, $p<0.00038$. Locke also computed the correlation between his obtained ranks and ranks based on children’s order of acquisition derived from a study by Templin (1957). He found a moderate, positive correlation (.49). When we correlated our rankings with the child acquisition rankings, we obtained a high positive correlation: $r_s=.71$, $t(18)=4.3$, $p<.0002$.

4. Discussion

We obtained a moderate, but not a high correlation between the ratings of our participants and the ratings of Locke’s participants for this task. Our data were more highly correlated with child order of acquisition than were Locke’s, supporting his interpretation that children acquire sounds that adults rate as easier at a younger age and sounds that adults rate as being more difficult at an older age. Our data reported in Table 1, column 4 differ from Locke’s in

column 2 in that we reported whole numbers. This is because we obtained whole numbers when we followed his procedure: “Data were tabulated (Table 1) by tallying the number of times each phoneme was encircled by the total group of 55 subjects (p. 195).” It appears that he performed an additional calculation that was not described in the study, because he reported fractional values in column 2. However, our analysis was based on the obtained ranks (columns 3 and 5), so the scaling of those values should not matter.

Replication Experiment 2

5. Material and methods

5.1 Participants—The participants were 50 undergraduate students. The inclusion/exclusion criteria were the same as for Experiment 1.

5.2 Materials—We created 53 written lists of the 20 phonemes followed by “uh” (e.g., 20 syllables - puh, nuh, fuh, etc). There was a scale from 1–9 beside each syllable. The syllable order was different for each list.

5.3 Procedure—The lists were handed out in a senior-level course in speech-language pathology. Each participant received a different list. The directions were the same as those used by Locke:

“Your job is to whisper each sound and try to decide how much effort it takes to say it. ‘Effort’ means the amount of muscular tension or strain in the lips, jaw, tongue, or throat needed to whisper the sound. ‘Little effort’ means a sound seems to require little muscular tension or strain in the lips, jaw, tongue, or throat. ‘Great effort’ means a sound seems to require great muscular tension or strain in the lips, jaw, tongue, or throat. If a sound seems to require little effort, encircle the numbers 1 or 2. If a sound seems to require great effort, encircle the numbers 8 or 9 (p. 195–196).”

Again, participants were told that if they were non-native speakers of English, they should put a ‘1’ at the top of their paper, and if they had a history of speech and/or language problems or therapy, they should put a ‘2’ at the top of the paper. Following Locke’s procedure, we calculated the average rating for each phoneme to get an overall rating of effort for that phoneme. We then used a Spearman rank order correlation to compare our ratings with Locke’s. As in Experiment 1, Locke computed the correlation between his obtained ranks and the children’s order of acquisition derived from Templin’s study (1957), and we did this for our data, as well.

6. Results

Four participants were excluded who had previous speech therapy, and one was excluded for circling more than one number for several phonemes. Columns 6 and 7 in table 1 show Locke’s ease ratings and the corresponding phoneme rankings. Columns 8 and 9 in table 1 show our ease ratings and the corresponding phoneme rankings. The statistical analysis revealed a very high, significant positive correlation between our phoneme ranks and Locke’s: $r_s = .93$, $t(18)=10.73$, $p<.000001$. Locke’s ranks were moderately positively

correlated with child order of acquisition (.66), while our data were highly positively correlated with child order of acquisition: $r_s=.74$, $t(18)=4.61$, $p<.0001$. We compared the correlation between the data from Experiments 1 and 2 for our data and for Locke's data in order to further explore replicability. He reported a high positive correlation, .78, while we obtained a higher positive correlation between the two methods: $r_s=.86$, $t(18)=7.4$, $p=.000001$.

7. Discussion

The task of rating ease of articulation is fairly subjective, even for mature speakers of the language. However, the nearly perfect positive correlation between our phoneme rankings and Locke's rankings suggests that the method used in Experiment 2, i.e., rating ease on a 9-point scale, is quite robust, despite the subjective nature of the task.

8. General Discussion

We were able to replicate the results of Locke's experiments, with especially close replication of Experiment 2. According to the data from Makel et al. (2012), we are in a minority. As noted earlier, direct replications by investigators who were not associated with the original study are likely to result in a failure to replicate the original results. There are a variety of reasons for direct replication failures. These include a lack of detail regarding methods for conducting the experiment and/or for analyzing the data and selective reporting of the original data. One aspect of Locke's study that we were not able to replicate was the nature of the participants. His were from chemistry, basic speech and German classes for Experiment 1 and basic speech and chemistry courses for Experiment 2, however, the number of students from each class were not provided, therefore, we could not duplicate the nature of the participants.

Our participants were majors in speech-language pathology and audiology, and thus may have known more about articulation than Locke's participants. This may be one reason why we had a lower correlation between our data for Experiment 1 and his data for Experiment 1. However, it is likely that the students taking German would have been instructed in pronunciation of that language and how it differs from English, and, depending on the nature of the basic speech class, diction might have been addressed in that class. In addition, if the amount of knowledge and background in speech production was the important factor, we should have obtained a lower correlation for Experiment 2, where our subjects were seniors, than for Experiment 1, where our subjects were juniors. We did obtain a higher correlation between the two methods for rating ease than Locke did, so perhaps this is because the participants in both studies were majors in speech-language pathology and audiology. Yet that correlation was still not as high as between our data for Experiment 2 and his data for Experiment 2.

An important question is why we were able to replicate Locke's results, especially so closely for Experiment 2, given that this outcome is reportedly rare. One reason is that he provided sufficient information regarding the procedure and data analysis, so that we were able to follow it exactly, with the exception of participant selection (described above). Another reason may be that he developed experimental tasks that have good reliability and validity.

A third reason may be that the study requires no subjectivity on the part of the investigators. We needed only to tally the numbers and perform the statistical analysis. Investigators may want to consider these factors when designing and reporting a study, so that a successful replication, if undertaken, is more likely.

As noted earlier, some investigators believe that the replicability crisis has been exaggerated and that there are better ways to evaluate the reliability and validity of scientific findings than through the use of direct replication, such as by employing meta-analyses (e.g., de Winter & Happee, 2014; Stanley & Spence, 2013). However, the creation of the Reproducibility Initiative, the new initiatives announced by the U.S. NIH, and the papers published in high profile scholarly journals on the issue of reproducibility suggest that many investigators do not believe that the issue is exaggerated. It is clear that not every experimental finding needs to be directly replicated. How do we determine which studies should be replicated? The Reproducibility Initiative is targeting high profile papers. In psychology, some investigators are targeting high profile phenomena (those for which many papers have been published and/or cited). One high profile phenomenon that independent investigators in psychology have selected for direct replication is the goal priming effect. In the original study, Bargh, Chen, and Burrows (1996) found that when participants were exposed to stereotypes of aging, they walked more slowly when leaving the lab, and this phenomenon was labeled the goal priming effect. Several investigators have been unable to replicate either the original study or a variety of subsequent goal priming studies (Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, Pashler, 2013; Pashler, Coburn, & Harris, 2012). Interestingly, Doyen and colleagues were only able to replicate the findings of the original study when *the investigators* were led to believe that the participants would walk more slowly. The two studies that Harris et al. failed to replicate were reported in a paper that has been cited more than 1400 times (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001).

Which findings are worthy of direct replication in communication disorders? Muma (1993) suggested that the need for replication of the research on some subpopulations of individuals with communication disorders might be greater than for other subpopulations. One important category, however, is treatment studies (Muma, 1993; Onslow, 1992). Well-documented successful and unsuccessful direct replications of treatment studies with careful descriptions of the participants could provide extremely valuable data regarding the factors that influence treatment outcomes. In addition, studies that are aimed at providing support for or refuting theories of the mechanisms that underlie a particular communication disorder also are worthy of direct replication. An example would be investigations of the nature of speech errors (phonemic or phonetic) to distinguish speech problems due to aphasia/apraxia of speech. We felt that the Locke (1972) study was worthy of replication, because it was designed to investigate factors that underlie patterns of speech sound acquisition in children learning American English. Data regarding these factors can help inform models and theories of both normal speech sound acquisition and speech sound disorders. In addition, it was a unique method for getting insight into the relative difficulty of producing the various speech sounds of American English that can complement more objective measures, such as electromyography or strain gauge measures. In addition, if the insight of mature speakers of

the language is reliable and replicable, as these results suggest, perhaps it could be tapped for other purposes. For example, some of the techniques that we employ for the remediation of communication disorders are effortful, in that (at least in the early stages) they require the speaker to be constantly and consciously vigilant regarding *how* speech is being produced (e.g., slowed rate, easy onsets of articulation). It appears to be the case that some approaches require more perceived effort than others, and that could be a factor in selecting among different treatments. In the area of stuttering treatment, Ingham and colleagues (2009; Ingham, Warner, Byrd, & Cotton, 2006) have found that both persons who stutter and non-stuttering controls rate different fluency-inducing techniques as requiring different amounts of effort. Perhaps some individuals are better suited for more effortful approaches than others, and this may help explain differences in response to treatment.

Both our data and Locke's data revealed a correlation between ratings of motor ease by adults and order of speech sound acquisition in children. It is clear, however, that motor ease is not the sole factor underlying speech sound acquisition, nor is motor ability alone sufficient for acquiring normal speech. There must be an interaction between the motor system and the perceptual and cognitive-linguistic systems. Infants and children must have the motor ability to produce (increasingly) accurate speech, the perceptual ability to extract the sound patterns from a running stream of speech, and the cognitive ability to constantly compare one's own output with the output of others, until the two converge (Kuhl, 1987). Research has shown that there are both universal trends and individual variability in the particular order in which children acquire speech sounds in American English and other languages. In addition to motoric, perceptual and cognitive-linguistic factors, the frequency with which sounds occur in a particular language also appears to play a role in the more universal patterns of relative order of mastery, although it cannot completely account for them (Beckman & Edwards, 2010; Edwards, Beckman, & Munson, 2015). Thus, there are multiple factors that influence the child's ability to master the sounds of the language. Individual differences in ability may determine which factor influences a given child's order of mastery, as well as the nature of a particular child's speech sound disorder.

9. Conclusions

We were able to replicate the findings from Locke's study. Moreover, we found a high correlation between the two methods for rating ease of articulation that was higher than Locke reported in the original article. Both of the rating methods are subjective, so the very high positive correlation for the motor ease rating task between Locke's data and our data for Experiment 2 and the high correlation between the two methods for our data is especially striking. The data support Locke's idea that adult speakers, through explicit introspection, can successfully evaluate a linguistic skill that they acquired through implicit learning. Our replication of his findings suggests that these ease rankings have validity and could be considered in the analysis and treatment of speech sound errors in children and adults, at least for those who are native speakers of American English. An important caveat is that these are relative rankings. That is, sounds within the phonetic inventory of American English were ranked *relative to one another*. It is probably unlikely that these same consonant rankings would be obtained from speakers of other languages with different phonetic inventories. This may be especially true for tone languages, where more complex

laryngeal articulations, in addition to the supraglottic articulations, are required to signal meaningful lexical differences (Xu, 2004). Thus, the role of ease of articulation (as opposed to other factors, such as perceptual difficulty or frequency of occurrence of the sound in a language) and order of speech sound acquisition in other languages remains to be explored.

Acknowledgements

The second author was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under grant number R15DC011136 to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Bakker M, Van Dijk A, Wicherts JM. The rules of the game called psychological science. *Perspectives on Psychological Science*. 2012; 7(6):543–554.
2. Bargh JA, Gollwitzer PM, Lee-Chai A, Barndollar K, Trötschel R. The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology*. 2001; 81(6):1014–1027. [PubMed: 11761304]
3. Bargh JA, Chen M, Burrows L. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*. 1996; 71(2):230–244. [PubMed: 8765481]
4. Beckman ME, Edwards J. Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology*. 2010; 1(2):319–343. [PubMed: 21113388]
5. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012; 483(7391):531–533. [PubMed: 22460880]
6. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014; 505(7485):612–613. [PubMed: 24482835]
7. De Winter J, Happee R. Why selective publication of statistically significant results can be effective. *PLoS one*. 2013; 8(6):e66463. [PubMed: 23840479]
8. Doyen S, Klein O, Pichon CL, Cleeremans A. Behavioral priming: it's all in the mind, but whose mind? *PLoS one*. 2012; 7(1):e29081. [PubMed: 22279526]
9. Edwards J, Beckman ME, Munson B. Frequency effects in phonological acquisition. *Journal of Child Language*. 2015; 42:306–311. [PubMed: 25644416]
10. Harris CR, Coburn N, Rohrer D, Pashler H. Two failures to replicate high-performance- goal priming effects. *PLoS one*. 2013; 8(8):e72467. [PubMed: 23977304]
11. Hinkle, D.; Wiersma, W.; Jurs, S. *Applied statistics for the behavioral sciences*. Chicago, IL: Rand McNally College Publishing; 1979.
12. Ingham RJ, Bothe AK, Jang E, Yates L, Cotton J, Seybold I. Measurement of speech effort during fluency-inducing conditions in adults who do and do not stutter. *Journal of Speech, Language, and Hearing Research*. 2009; 52(5):1286–1301.
13. Ingham RJ, Warner A, Byrd A, Cotton J. Speech effort measurement and stuttering: investigating the chorus reading effect. *Journal of Speech-Language-Hearing Research*. 2006; 49(3):660–670.
14. Ioannidis JP. Why most published research findings are false. *PLoS medicine*. 2005; 2(8):e124. [PubMed: 16060722]
15. Ioannidis JPA. Why science is not necessarily self-correcting. *Perspectives on Psychological Science*. 2012; 7(6):645–654.
16. Kent, RD. *The speech sciences*. San Diego: Singular Publishing Group, Inc.; 1997.
17. Kent RD. The uniqueness of speech among motor systems. *Clinical Linguistics and Phonetics*. 2004; 18:495–505. [PubMed: 15573486]
18. Kuhl, PK. *Handbook of infant perception*. New York: Academic Press; 1987. Perception of speech and sound in early infancy. In; p. 275–382.

19. Locke JL. Ease of articulation. *Journal of speech and hearing research*. 1972; 15(1):194–200. [PubMed: 5012806]
20. Makel MC, Plucker JA, Hegarty B. Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*. 2012; 7(6):537–542.
21. Muma JR. The need for replication. *Journal of Speech and Hearing Research*. 1993; 36:927–930. [PubMed: 8246481]
22. Onslow M. Choosing a treatment procedure for early stuttering: issue and future directions. *Journal of Speech and Hearing Research*. 1992; 35:983–993. [PubMed: 1447932]
23. Pashler H, Coburn N, Harris CR. Priming of social distance? Failure to replicate effects on social and food judgments. *PloS one*. 2012; 7(8):e42510. [PubMed: 22952597]
24. Pashler H, Harris CR. Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*. 2012; 7(6):531–536.
25. Pashler H, Wagenmakers E. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*. 2012; 7(6): 528–530.
26. Stanley DJ, Spence JR. Expectations for replications: Are yours realistic? *Perspectives on Psychological Science*. 2014; 9(3):305–318.
27. Templin, MC. *Certain Language Skills in Children*. Minneapolis, MN: University of Minnesota; 1957. Institute of Child Welfare Monograph No. 26.
28. Xu Y. Understanding tone from the perspective of production and perception. *Language and Linguistics*. 2004; 5:757–797.

Continuing Education Questions

1. According to Ioannidis (2005), the problem with the .05 probability level for statistical significance is that:
 - a. it is too high, thus not rigorous enough.
 - b. it is not used in many studies in communication disorders.
 - c. it does not represent the number of literature-wide false positives.
 - d. it is not applicable to some research designs.
2. The data suggest that this type of replication is significantly less likely to replicate the results of the original study.
 - a. A direct replication by investigators who were not involved in the original study.
 - b. A direct replication by the original investigator(s).
 - c. A conceptual replication by investigators who were not involved in the original study.
 - d. A conceptual replication by the original investigator(s).
3. Impediments to self-correction in the psychological sciences identified by Ioannidis (2012) include:
 - a. insufficient time for manuscript reviews.
 - b. difficulties in getting negative results published.
 - c. investigators moving on to different projects.
 - d. inadequate descriptions of experimental methods.
4. The method of rating ease of articulation using a rating scale from 1–9:
 - a. resulted in a lower correlation between the original study and the current study than did the method of comparing pairs of sounds.
 - b. resulted in a higher correlation between the original study and the current study than did the method of comparing pairs of sounds.
 - c. resulted in a perfect positive correlation between the original and current study.
 - d. resulted in a very low correlation between the original and current study.
5. The U.S. NIH, along with several scholarly journals,:
 - a. has created the Reproducibility Initiative.
 - b. believe the replicability crisis is exaggerated.
 - c. have taken no interest in the issue of replicability.

d. are encouraging authors of low profile research papers to allow independent investigators to replicate their findings.

Answer Key: 1. c; 2. a; 3. b; 4. b; 5. a

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Motor ease of articulation for word initial phonemes as determined by paired comparisons (columns 2, 3 Locke's Data; columns 4, 5 our data) and motor ease ratings on a scale of 1–9 (columns 6, 7 Locke's data; columns 8, 9 our data).

Table 1

Word-Initial Phoneme	Locke Motor Ease Paired Comparison	Locke Rank	Our Motor Ease Paired Comparison	Our Rank	Locke Motor Ease Rating Scale	Locke Rank	Our Motor Ease Rating Scale	Our Rank
t	40.5	6	57	2	4.288	8	2.9	3.5
d	54.5	1	78	4	3.735	2	2.8	2
n	45.5	4	70	3	3.788	4	3.5	9
b	25.0	13	91	10	4.226	5.5	2.9	5
m	39.0	9	87	9	4.301	9	3.3	7
w	12.0	18	124	17	4.773	11	4.2	13
h	46.0	3	49	1	2.603	1	2.5	1
p	40.0	7.5	82	6.5	4.735	10	3.3	8
k	24.0	14.5	81	5	4.886	12	3.8	10
f	38.5	10	105	11	4.283	7	3.8	11
g	24.0	14.5	112	13	5.622	18	4.6	16
s	49.0	2	83	8	3.754	3	2.9	3.5
j	34.5	11	106	12	4.981	14.5	4.0	12
l	40.0	7.5	82	6.5	4.226	5.5	3.1	6
r	28.5	12	132	19	5.283	16	4.7	17
dʒ	9.0	19.5	123	16	5.698	19	4.9	18.5
ʃ	43.5	5	125	18	4.981	14.5	4.5	15
tʃ	9.0	19.5	141	20	6.207	20	5.0	20
z	20.5	17	122	15	5.390	17	4.9	18.5
v	22.5	16	121	14	4.943	13	4.5	14