# Letter to the Editor

## Response to Saltuklaroglu, Kalinowski, and Stuart (2010)

**Purpose:** To reply to the criticisms of Saltuklaroglu, Kalinowski, and Stuart (2010) by addressing their concerns regarding our study's methodology, statistical analyses, and findings. Also, to challenge what we view as omissions, misinterpretations, and inaccuracies on their part.
**Results:** Our operational definition of stuttering was sound. Participant adherence to the treatment protocol was telling and appropriately enforced. The question-asking task was proper given participant characteristics. Statistical analyses of treatment effects were correctly interpreted. Our general conclusions regarding the clinical merit of the SpeechEasy were misinterpreted by Saltuklaroglu and colleagues; our findings were in fact far less nullifying and more balanced than what they claim.
**Conclusions:** While robust immediate effects of altered auditory feedback (AAF) in the laboratory are well documented, recent longitudinal experiments conducted in naturalistic settings have found less consistent and pronounced effects with the SpeechEasy. These reports also indicate that initial reductions in stuttering are often not maintained over time. Future efforts to determine why this is so would be worthwhile.

**KEY WORDS: stuttering, treatment, altered auditory feedback, clinical trials, prosthetic devices**

Saltuklaroglu, Kalinowski, and Stuart (2010) begin by stating that "claiming their data represented Phase I clinical evidence, Pollard et al. concluded that the SpeechEasy was not therapeutically useful" (p. 908). Our study (Pollard, Ellis, Finan, & Ramig, 2009) was an appropriate Phase I clinical trial by any interpretation of the phrase. Using a sufficient sample size, we sought preliminary evidence of population effects and utilized a protocol allowing us to identify whether any individuals showed an unusual response against the group trend (Robey, 2004). We did not find evidence of a population effect across four months' time. However, rather than concluding that the device was not therapeutically useful, as Saltuklaroglu et al. erroneously claim, we instead stated that the general thrust of our results, taken together, indicated the device would likely be beneficial for some clients.

### *Response to Questions on Methodology*

Saltuklaroglu et al.'s (2010) first methodological concern pertains to stuttering counts. They believe that our operational definition of stuttering "is confounding for a number of reasons. First, starters and fillers (or interjections) are typically not considered signature stuttering events as they can also be found in the 'disfluent' speech of normal speakers" (p. 909). Here, they omit the rest of our definition in which we explained that "a starter or filler was counted as a disfluency only if it was determined that

the participant used it habitually to postpone the next word or as a means to say the desired word fluently" (Pollard et al., 2009, p. 521). Our high interrater reliability for stuttering counts indicated that determinations regarding the fluency status of interjections, and all other types of disfluencies for that matter, were quite reliable. The definition of stuttering to which Saltuklaroglu et al. object was taken from a well-regarded textbook on stuttering (Guitar, 2005). We chose it because it captures all that the average listener would consider to be pathological disfluencies. To illustrate, the speech of normally fluent speakers does include numerous interjections, but it rarely if ever includes what most would consider to be a signature feature of stuttering: using interjections conspicuously in the manner described above. For instance, a person who uses the interjection "um" preventively, saying "Um, um, um, um, I want to go home" cannot be said to have spoken that sentence fluently, despite not stuttering on any word in the phrase "I want to go home." By Saltuklaroglu et al.'s rationale, no part of that utterance should be counted as disfluent because "um" is an interjection that appears in the speech of normally fluent people. Relatedly, the speech of normally fluent speakers also includes frequent whole word and phrase repetitions, none of which would be considered stuttering behaviors in the pathological sense. Are we not to count those types of disfluencies as well in individuals who stutter?

Saltuklaroglu et al. (2010) state that we

> asked participants to produce volitional initiating gestures that, using their definition, would then be counted as stuttering behaviors. Simply put, by using this definition, true stuttering events could not be adequately separated from 'normal disfluency' or from the motor strategies prescribed by the treatment protocol. (p. 909)

First, we did not ask our participants to produce starter sounds. Rather, we taught the techniques to them (as indicated in the SpeechEasy protocol), informing them that "these active strategies could be introduced at their discretion to help initiate voicing and/or enhance responsiveness to the second speech signal and could be reduced if they felt choral effects and naturalness stabilize" (p. 520). More to the point, we wished to evaluate the entire SpeechEasy protocol, which included both (1) use of the device and (2) teaching of elective techniques that, presumably, enhance responsiveness to the device. We could not have omitted any part of the protocol because we wanted to examine the device as it was intended to be used. Any initiating gestures that prolonged a sound would be counted as disfluent, and rightfully so, since sound prolongations are in fact a type of disfluency associated with stuttering. Furthermore, attempting to distinguish "true stuttering" from "volitionally produced initiating gestures" would have been a speculative exercise in telepathy. If a phoneme, syllable, or word sounds disfluent, it is disfluent, regardless of the inscrutable intentions of the person uttering it. However, this is a moot point because our interrater reliability data demonstrated excellent agreement between scorers.

Saltuklaroglu et al. (2010) maintain that "counting volitionally produced motor techniques as disfluencies would further exacerbate stuttering counts during the treatment phase and inherently minimize or negate any treatment effect" (p. 909). We cannot agree more. If faithful implementation of the SpeechEasy protocol potentially resulted in an increase in disfluencies, as they maintain may have occurred, then surely that is an effect of the treatment that ought to have been measured in the treatment phase.

Regarding scoring of the Stuttering Severity Instrument–Third Edition (SSI-3), our results showed that scores on this instrument either improved or remained unchanged for all participants at the beginning and end of the treatment phase compared to baseline. Indeed, this was one piece of evidence supporting our contention—reiterated throughout the paper—that the SpeechEasy may be clinically useful for some clients.

Saltuklaroglu et al. (2010) argue that "external validity is compromised as Pollard et al. cannot compare their findings with other studies evaluating treatment effects of the SpeechEasy" (p. 909). External validity does not involve the ease with which one study can be compared to another, but rather the ease with which the findings of a single study can be generalized to a larger population. In that connection, we will restate here what was stated in our paper: "We feel that the combination of less restrictive inclusion criteria and repeated, naturalistic sampling procedures increases the external validity of our results" (p. 530). We tested the effects of the SpeechEasy under more challenging conditions and with a broader range of PWS than have commonly been employed in past AAF research. Rather than compromising our external validity, our methods increased it substantially.

Saltuklaroglu et al.'s (2010) next methodological concern pertains to device wear-time. They ask "How can one determine the efficacy of the treatment in a Phase I clinical trial without adherence to a treatment protocol where usage patterns were highly variable" (p. 909)? We would argue the contrary point, which it appears they failed to appreciate: namely, perhaps usage patterns were highly variable *because* of the efficacy of the treatment. Consider that the average wear-time for the group was 5.0 hr per day, meaning that the group did in fact meet the target wear-time. Of the five participants who wore the device less than 5 hr per day, three of them wore it between 4.0 and 4.4 hr per day. It is reasonable to assume that a difference as meager as 24–60 min per day would have minimal impact on stuttering frequency. But that is not the main problem with their criticism. The main problem

is their contention that "variable changes in stuttering frequency might simply reflect the variable compliance to the protocol, leading one to question if better compliance would produce a stronger treatment effect" (p. 909). We would argue that they are inverting cause and effect. That is, a stronger treatment effect would have led to better compliance. Our self-report data support this interpretation. The two participants with the lowest wear-times also had the most negative opinions of the device; one of them returned it after a month because it was too bothersome to continue using. That, in itself, was a clear treatment effect, but apparently not one that Saltuklaroglu et al. would consider valid. Were we to force that participant to continue using a device that she detested? Judging by their logic, it would seem so. They lament that the treatment protocol was "not enforced," but how could we have ethically forced our participants to use an experimental treatment more frequently than they wished? Moreover, we sought real-world usage data on the SpeechEasy, which would have been confounded by coercion from the investigators.

Saltuklaroglu et al.'s (2010) objection to our question-asking task again demonstrates their unilateral reasoning. They state that "any stuttering event that is recorded in such a small speech sample will disproportionately inflate the overall percentage of stuttered syllables" (p. 909). What they fail to acknowledge is, again, the contrary: A small speech sample will disproportionately inflate the overall percentage of fluent syllables as well. To illustrate, eight treatment phase speech samples during this task were scored as perfectly fluent (i.e., 0% stuttered syllables). By contrast, none of the treatment phase speech samples from the 300 syllable-long conversation task were perfectly fluent. Rather than underestimating the reduction in stuttering during the question task, our modest sample size may just as likely have overestimated it, increasing the likelihood of obtaining a treatment effect. However, another ethical issue needs to be raised here. We fully acknowledged the relatively small sample size for this task, but

> felt that requiring participants to ask several questions to many different strangers would have been too daunting or, in some cases, an impossible expectation. Some participants had little or no experience with desensitization exercises commonly used to lessen the fear and avoidance of stuttering, making merely one question a sufficiently stressful undertaking. (p. 521)

It would have been unprincipled of us to have demanded more questions, potentially subjecting our participants to undue psychological or emotional stress.

## Response to Questions on Statistical Analyses

Regarding our statistical analyses, both Bartlett's test ($p = .60$; Snedecor & Cochran, 1983) and Levene's test ($p = .89$; Levene, 1960) failed to show heteroscedasticity in our pooled data. Thus, the assumption of sphericity was not violated. However, we appreciate Saltuklaroglu et al.'s (2010) point regarding the positive skewness of %SS data. We therefore used the Kolmogorov-Smirnov test (Chakravart, Laha, & Roy, 1967) as a goodness of fit normality test for each task (i.e., reading, conversation, question) by phase (i.e., baseline, treatment, withdrawal) distribution. The tests revealed that, at the individual level, our distributions were non-Gaussian. We then used a Friedman nonparametric repeated measures analysis of variance (ANOVA) to test for differences between experimental phases for each speech task (Delorme, 2006). All results failed to reach significance, indicating no treatment effect for any speech task: for reading, $\chi^2(26) = 2.03$, $p > .05$; for conversation, $\chi^2(26) = 1.27$, $p > .05$; for question, $\chi^2(26) = 0.03$, $p > .05$. The effect size as calculated by Cohen's $d$ indicated a small to medium effect size for experimental phase ($d = .43$, 95% CI = .18–.68; Cohen, 1988).

## Response to Questions of Interpretations

Lastly, Saltuklaroglu et al. (2010) question our interpretations, saying that "though the authors addressed the positive qualitative data in their discussion as 'intriguing' and 'curious,' little if any importance was attributed" (p. 910). This appears to be another example of selective reading. We described the discrepancy between fluency counts and subjective impressions as "something that is beginning to emerge as an intriguing phenomenon in SpeechEasy research" (p. 528). We then provided three examples of similar findings in the literature (Cook & Smith, 2006; Molt, 2006; Runyan, Runyan, & Hibbard, 2006). In the same paragraph, we stated that, "60% of our sample mentioned increased confidence in speaking as a benefit of device use" (p. 528). The same studies were again cited as further evidence to suggest that increased confidence may be involved with the phenomenon in question.

As for the degree of importance we attributed to positive qualitative data, we feel that Saltuklaroglu et al. (2010) have grossly misinterpreted the crux of our paper. We allotted a sizeable portion of our Discussion section to explaining the implications of the different types of positive qualitative data we collected. This was done to highlight the device's effectiveness with certain participants, despite its failure to produce empirical effects at the group level. Our summative statement of the device's effectiveness—which drew from our entire dataset—was much more nuanced and balanced than Saltuklaroglu et al. acknowledge. It reads as follows:

> Our group findings showing no treatment effect for the device suggest that Phase II trials of the SpeechEasy are not warranted; however, given the parameters of participant recruitment and methodological protocols

> employed in the study, this is a qualified conclusion. As some participants benefited clinically on certain speech tasks and/or reported subjective satisfaction with the device, clinicians may wish to probe for device effects with clients on an individual basis. (p. 530)

This is far from a blanket statement that the device is worthless and the door ought to be closed on future explorations of its clinical worth. It certainly runs counter to Saltuklaroglu et al.'s claim that we "concluded that the device was not therapeutically useful and further testing is unwarranted" (p. 908).

Regarding Saltuklaroglu et al.'s (2010) contention that we should have included additional outcome measures, we concur that duration of stuttered events and speech naturalness ratings are useful and we encourage future research on the SpeechEasy to incorporate such measures. Our reason for not doing so was logistical. The research protocol included a great deal of objective and subjective data, the subjective measures in fact comprising the bulk of our findings. We simply did not have space in the manuscript to include an exhaustive array of outcome measures. Also, we did acknowledge this limitation of our study, saying that "there may have been unaccounted-for variables that could have determined participants' ultimate opinions of the device" (p. 528).

We take issue with the citations used to support Saltuklaroglu et al.'s (2010) conclusion that "there has been replication of positive group effects with the SpeechEasy device, albeit variability across individuals (Armson & Kiefte, 2008; Armson et al., 2006; Kalinowski, 2003; Kalinowski et al., 2004; O'Donnell et al., 2008; Stuart et al., 2004, 2006)" (p. 910). Here, one must inspect the details of this claim. Kalinowski (2003) was an "autobiographical clinical commentary" from one of the SpeechEasy's developers as he wore the device while teaching. It did not report group effects. Armson and Kiefte (2008) and Armson et al. (2006) both reported significant immediate group effects of the device across speech tasks, but did not report longitudinal data. Incidentally, our study reported similar findings upon application of the device, further strengthening the literature on immediate group effects of AAF. O'Donnell et al. (2008) used a multiple single-subject design and did not perform statistical analyses of group data. Interestingly, they employed a longitudinal design similar to ours and obtained speech samples in extraclinical settings as well. They reported that, "at the end of the study, while four of the seven participants experienced a reduction in stuttering when wearing the device, three exhibited more stuttering with the device than without. The net effect was a group mean reduction of 14%" (p. 112). Stuart et al. (2004, 2006) was actually a single study: The same cohort of eight participants was followed up at 4 months (2004) and 12 months (2006) postfitting using the same outcome measures. Kalinowski et al. (2004) collected surveys from individuals who had already liked the SpeechEasy well enough to purchase it. Thus, upon further scrutiny, we are left with only two studies—both, coincidentally, conducted by the developers of the SpeechEasy—as the only longitudinal evidence of positive group effects for the device. Two studies simply does not constitute "a robust corpus of contradictory data" claimed by Saltuklaroglu et al. (2010, p. 908).

Contrary to what Saltuklaroglu et al. (2010) maintain, there is not currently a "corpus of predominantly positive group effects" (p. 910) for the device, and certainly not in extraclinical settings, which is of course where those who stutter live their lives. Instead, there are two longitudinal experiments conducted in naturalistic environments (O'Donnell et al., 2008; Pollard et al., 2009). The results from each suggest that the SpeechEasy is beneficial for some and not for others, and that immediate group reductions in stuttering are not maintained over several months. Determining the reason(s) for both of those findings is a worthwhile endeavor for future researchers. We hope to see efforts in that direction.

Ryan Pollard
University of Colorado at Boulder

Don Finan
University of Northern Colorado, Greeley

Peter R. Ramig
University of Colorado at Boulder

## Acknowledgment

We thank Phillip Gilley for helpful suggestions regarding statistical methods.

## References

**Armson, J., & Kiefte, M.** (2008). The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. *Journal of Fluency Disorders, 33,* 120–134.

**Armson, J., Kiefte, M., Mason, J., & De Croos, D.** (2006). The effect of SpeechEasy on stuttering frequency in laboratory conditions. *Journal of Fluency Disorders, 31,* 137–152.

**Chakravarti, I. M., Laha, R. G., & Roy, J.** (1967). *Handbook of methods of applied statistics* (Vol. 1). New York, NY: Wiley.

**Cohen, J.** (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

**Cook, M. J., & Smith, L. M.** (2006, November). *Outcomes for adult males using the SpeechEasy fluency device for one year*. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, Miami, FL.

**Delorme, A.** (2006). Statistical methods. *Encyclopedia of Medical Devices and Instrumentation, 6,* 240–264.

**Guitar, B.** (2005). *Stuttering: An integrated approach to its nature and treatment* (3rd ed.). Baltimore, MD: Williams & Wilkins.

**Kalinowski, J.** (2003). Self-reported efficacy of an all in-the-ear-canal prosthetic device to inhibit stuttering during one hundred hours of university teaching: An autobiographical clinical commentary. *Disability and Rehabilitation, 25,* 107–111.

**Kalinowski, J., Guntupalli, V., Stuart, A., & Saltuklaroglu, T.** (2004). Self-reported efficacy of an ear-level prosthetic device that delivers altered auditory feedback for the management of stuttering. *International Journal of Rehabilitation Research, 27,* 167–170.

**Levene, H.** (1960). Robust tests for equality of variances. In I. Olkin et al. (Ed.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford, CA: Stanford University Press.

**Molt, L.** (2006, November). *SpeechEasy AAF device long-term clinical trial: Attitudinal/perceptual measures*. Poster presented at the Annual Convention of the American Speech-Language-Hearing Associating, Miami, FL.

**O'Donnell, J. J., Armson, J., & Kiefte, M.** (2008). The effectiveness of SpeechEasy during situations of daily living. *Journal of Fluency Disorders, 33,* 99–119.

**Pollard, R., Ellis, J. B., Finan, D., & Ramig, P. R.** (2009). Effects of the SpeechEasy on objective and perceived aspects of stuttering: A 6-month, Phase I clinical trial in naturalistic environments. *Journal of Speech, Language, and Hearing Research, 52,* 516–533.

**Robey, R. R.** (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37,* 401–411.

**Runyan, C. M., Runyan, S. E., & Hibbard, S.** (2006, November). *The Speech Easy device: A three year study.* Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, Miami, FL.

**Saltuklaroglu, T., Kalinowski, J., & Stuart, A.** (2010). Refutation of a therapeutic alternative? A reply to Pollard, Ellis, Finan, and Ramig (2009). *Journal of Speech, Language, and Hearing Research, 53,* 908–911.

**Snedecor, G. W., & Cochran, W. G** (1983). *Statistical methods* (8th ed.). Ames, IA: Iowa State University Press.

**Stuart, A., Kalinowski, J., Rastatter, M., Saltuklaroglu, T., & Dayalu, V.** (2004). Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: Initial fitting and 4-month follow-up. *International Journal of Language and Communication Disorders, 39*(1), 93–113.

**Stuart, A., Kalinowski, J., Saltuklaroglu, T., & Guntupalli, V. K.** (2006). Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: One-year follow-up. *Disability and Rehabilitation, 28,* 757–765.

Received February 19, 2010

Accepted April 27, 2010

DOI: 10.1044/1092-4388(2010/10-0050)

Contact author: Ryan Pollard, University of Colorado at Boulder, Speech, Language & Hearing Sciences, 409 UCB, Boulder, CO 80309. E-mail: ryan.pollard@colorado.edu.