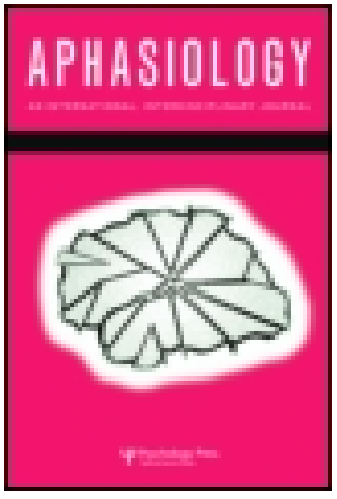

Aphasiology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/paph20

Establishing the effects of treatment for aphasia using single-subject-controlled experimental designs

Cynthia K. Thompson[bca]

[a] Aphasia and Neurolinguistics Research Laboratory, Center for the Neurobiology of Language Recovery, Evanston, IL, USA

[b] Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

[c] Department of Neurology, Northwestern University, Chicago, IL, USA

Published online: 24 Dec 2014.

Aphasiology, 2014
http://dx.doi.org/10.1080/02687038.2014.987043

COMMENTARY

Establishing the effects of treatment for aphasia using single-subject-controlled experimental designs

Cynthia K. Thompson*

Aphasia and Neurolinguistics Research Laboratory, Center for the Neurobiology of Language Recovery, Evanston, IL, USA; Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA; Department of Neurology, Northwestern University, Chicago, IL, USA

Single-subject-controlled experimental research methods were advanced several decades ago as an alternative to group experimental research in basic experimental psychology (Sidman, 1960) and subsequently adopted for examining the effects of treatment for a variety of disorders, including communication disorders (Kearns & Thompson, 1991; Kiran et al., 2013; McReynolds & Kearns, 1986; McReynolds & Thompson, 1986; Schlosser, 2003; Thompson, 2006; Thompson & Kearns, 1991). First utilised to study behavioural intervention primarily in children, these designs are now often used to study the effects of model-based (e.g., representational and processing models of language, including cognitive neuropsychological models) and other interventions for adults with aphasia.

Single-subject-controlled experimental research, like group experimental research, entails explicit requirements for demonstrating both *internal* and *external validity*. Internal validity or experimental control is essential to rule out placebo effects, Hawthorne effects,[1] and the influence other potential extraneous variables on the dependent measures employed in the study. In group research, this is accomplished using *experimental and control groups* of participants. The idea is that both participant groups are exposed to the same extraneous variables; therefore, when reliable between-group differences are found, this can be attributed to the treatment under study and not to some other variable(s). Group studies also use parametric statistics to address external validity, that is, generality of study findings to the population, although the assumptions of parametric statistics are often violated in aphasia treatment research.

In single-subject controlled experimental design, rather than employing experimental and control groups of participants, *experimental and control phases* are used, with experimental control (internal validity) established between phases rather than between groups. Importantly, the term *single subject* does not imply that only a "single subject" is studied. Rather, the term refers to the fact that using such designs, *experimental control* is demonstrated within individual subjects. One common single-subject-controlled experimental design used for studying aphasia treatment is the multiple baseline design across behaviours (i.e., sets of items). In this design the independent variable (treatment) is consecutively applied to one set of items at a time in experimental phases, following control phases as shown in Figure 1. Internal validity is forthcoming when changes in dependent measures (i.e., item sets) occur in experimental phases compared to control phases.

*Email: ckthom@northwestern.edu

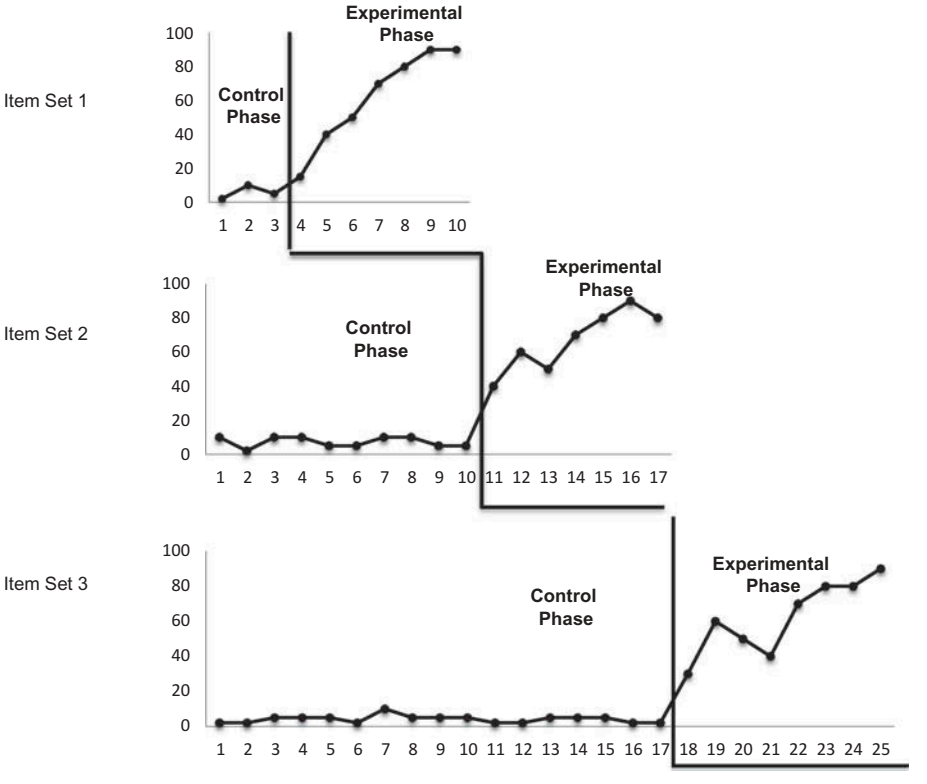

Figure 1. Illustration of a single-subject-controlled experimental design: the multiple baseline design across item sets, frequently used in aphasia treatment research. Hypothetical data are shown in control and experimental phases, arranged to allow demonstration of treatment effects, with internal validity, for one participant (performance accuracy (ordinate) for three items sets across sessions (absciss)). Note that for a multiple baseline design across item sets, additional participants are required for counterbalancing and replication (see text).

External validity in single-subject-controlled experimental design is addressed by *direct* and *systematic replication of treatment effects*. Using the multiple baseline design, training across sets of items allows direct, *within-participant replication*. This replication shows that the same treatment is effective for improving more than one set of items. Without this replication, it is possible that treatment is item specific (or item-set specific), which in and of itself is weak evidence of treatment effects. In addition, single-subject-controlled experimental designs require direct, *across-participant replication*. Using a multiple baseline design across three sets of items, for example, requires minimally three participants, with counterbalanced application of treatment across item sets (to control for order effects), resulting in nine potential replications of the treatment effect. Ideally, a total of six participants are included in a multiple baseline study with three item sets, further strengthening the design and allowing 18 possible replications. In addition, systematic replication in follow-up studies examining the effects of the same treatment for different participants, in different settings, and so on, further adds to external validity. As the number of replications increase, confidence that the treatment effect is generalisable to other people with aphasia, with language deficits similar to study participants, increases.

Howard, Best, and Nickels (2014) present a critical evaluation of single-subject-controlled experimental designs (termed Approach A) and suggest an alternative single-case study approach (Approach B), indicating that Approach B is the "way forward" for aphasia treatment research. Although all approaches to scientific inquiry are (and should be) subject to critical evaluation and, in turn, they evolve over time, reflecting advances in methods of behavioural observation and measurement, data management, and other influences, the points raised by Howard et al. (2014) reflect an incomplete understanding of the methods used in Approach A and the reasons for using them. In this paper, I detail some of the authors' misconceptions about Approach A and clarify the requirements for single-subject-controlled experimental research.

Approaches A and B: are they similar?

Howard et al. (2014) suggest that Approaches A and B are alike in that both usually include "three phases"—a pre-therapy or 'baseline' phase, a therapy phase, and a post-therapy phase (pp. xx–xx); however, the assumptions made about Approach A are not completely correct. The authors make the following points:

(1) In both the approaches a demonstrated lack of change in performance across pre-therapy probes followed by improvement with treatment is taken as evidence for the effectiveness of treatment.
(2) In the therapy phase there are usually at least two sets of items: a set that [is] treated and an untreated control set, and performance on the both sets is monitored.
(3) If both the treated set and the control set improve during therapy, this is evidence that generalisation has occurred, and if the treated set improves more than the control set, at least part of the improvement is item-specific.
(4) There may be more than one therapy phase: in a second therapy phase, the same treatment might be applied to a different set of items, the effects of another therapy method may be explored, or a different task or process may be treated.

I address each of these points, in turn. With regard to Point 1, the authors are correct that change in performance when treatment is applied is evidence of the effectiveness of treatment. However, the relation between baseline performance and that during treatment is *only partial evidence for effective treatment* for Approach A. That is, Approach A designs require replication of the treatment effect within subjects, for example, using a multiple baseline design across item sets as discussed earlier. Simply showing a change from baseline to treatment does not constitute experimental control or internal validity.

With regard to Point 2, inclusion of both treated and untreated control items, while this may be the case for Approach B, it is not for Approach A. As pointed out earlier, multiple baseline designs do include (and require) at least two (or three) item sets; however, they are not selected to be "treated" and "control" sets. Rather, treatment (experimental) and control phases are introduced for all the sets of items. In addition, there are many other designs within Approach A, including the multiple baseline design across participants or settings that do not require multiple item sets. Further, when item sets are monitored during the study, in Approach A performance on all sets is monitored continuously throughout both baseline and treatment phases of the study (see later for further discussion of repeated measurement requirements of Approach A). In Approach B, this typically occurs only at pre- and post-treatment.

The third point made by Howard et al. (2014) is also only true in part for Approach A —that is, when both treated and untreated sets improve, this reflects generalisation across items. For Approach A this pattern, in and of itself, constitutes a loss of experimental control. Using the multiple baseline design across item sets, for example, internal validity is lost if changes in the second (or third) set occur during treatment of the first (or second) set, since when/if this occurs there is no way to know whether or not extraneous variables affected performance. If generalisation is a possible, or expected outcome of treatment, additional design components are required (see Connell & Thompson, 1986). For example, a multiple baseline across participants can be added, allowing demonstration of experimental control when language improves for each participant, following baselines of increasing length. Because no other design components are included in Approach B, generalisation across sets also renders the results as internally invalid. When this happens it is simply not possible to rule out the effects of extraneous variables on changes in performance for either set.

Finally, although the authors do suggest, in Point 4, that both Approaches A and B may include more than one "therapy phase" (putatively akin to direct replication), with treatment applied to a different set of items, they also suggest that additional phases may be included to explore the effects of another treatment or a different task or process may be treated. Approach A, however, does not recommend application of different treatment approaches in adjacent study phases due to potential order effects. Rather, other designs are utilised for this purpose, for example, the alternating treatments design (Barlow & Hayes, 1979). In addition, application of treatment to different tasks or processes is not recommended because the treatment may not be appropriate for them, which could potentially preclude replication of treatment effects.

"Problems" with Approach A according to Howard et al.

Howard et al. (2014) also present what they suggest are "problems" with Approach A "features", including (1) the number of stimuli used and how they are selected, (2) baseline stability, (3) the length of the treatment phase(s), (4) repeated measurement of dependent variables, and (5) data analysis procedures. Although all of these are important to consider in aphasia treatment research, they are not features of single-subject-controlled experimental research, per se, with the exception of requirements for measurement of the dependent variables in the study ((2) and (4) given earlier). Rather, Approach A involves a number of other quality markers as listed in Table 1. Before turning to those, I discuss the points raised by Howard et al. (2014).

Experimental stimuli

Single-subject-controlled experimental design (Approach A) is silent with regard to the number of stimuli included in treatment studies and how to select them. Both are experimenter decisions based on the experimental questions under investigation. The authors criticise Approach A research for including what they consider to be small number of items and suggest, rather, that a large number of items are needed and are required in Approach B. Although more items may be better in some cases, for the reasons pointed out by Howard et al. (2014), a large number of items can also be disadvantageous. For example, if examining the effects of a novel treatment is the focus of the study, generation of a rapid and robust experimental effect may be desirable. If the treatment turns out to be ineffective, the use of many items to show this may be contraindicated. Interestingly, the hypothetical

Table 1. Single-subject-controlled experimental research quality indicators.

Dependent variables (baseline and treatment probe measures)
• Operationally specified
• Quantifiable
• Repeated measurement
○ Baseline probes establish pre-treatment patterns to predict pattern of future performance in the absence of treatment application
○ Treatment probes establish changes in performance patterns, compared to baseline
• Reliability of measurement (inter-observer agreement) obtained
Independent variable
• Operationally specified to allow replication of intervention methods in subsequent studies and/or in clinical practice
• Measurement of treatment fidelity
Internal validity (experimental control)
• Demonstrated change in dependent measures when (and only when) independent measures are introduced
• Inclusion of experimental and control phases to allow within participant replication of the treatment effect
• Inclusion of a sufficient number of participants for across participant replication and counterbalancing
External validity (generality of results)
• Direct replication of treatment effects within and across participants
• Systematic replication of treatment effects across studies

Note: Adapted from Gast (2010) and Thompson (2006).

data presented by Howard et al. (2014) show that treatment effects, in fact, are less robust when larger number of items are included in experiments. Thus, using a smaller set of items, within the context of a controlled experimental design, can provide initial evidence that a particular treatment does or does not improve language performance. Greater number of items may be included in follow-up replications, once initial treatment effects have been established. For example, in Thompson et al. (1997), one of our first single-subject-controlled studies investigating the effects of linguistically based treatment for sentence deficits (i.e., Treatment of Underlying Forms; TUF), we included a total of 60 items, with 15 items for the treatment of each of four sentence types. In a more recent study (Thompson, den Ouden, Bonakdarpour, Garibaldi, & Parrish, 2010), also investigating TUF, we used 20 items each to train three sentence types (total $n = 60$). Notably, in our current work, using a group experimental design to investigate the neurobiology of TUF our stimulus set includes 120 items.

Further, in some cases, a large number of items may be contraindicated. For example, studies examining treatment for sentence processing, requiring participants to produce multiple exemplars of various sentence types, which they are incapable of producing, may result in participant fatigue and/or frustration (and perhaps attrition). In addition, probing may require several hours. In our current treatment study examining the learning and generalisation effects of training long passive sentence structures, participants studied to date ($n = 9$) require up to 20 s to produce each response during baseline, when no feedback or experimenter assistance is provided. The point here is that Approach A researchers are free to, and do, include smaller or larger number of items depending on the domain of language being treated, the treatment under investigation, and the phase of experimentation. Hence, the number of items included in experiments is not a feature or requirement of single-subject-controlled experimentation.

Establishing baseline performance

Turning to the issue of pre-treatment performance, the authors indicate that whereas a feature of Approach A is to test "many times before treatment" begins (p. xx), in Approach B the length of baseline is "pre-determined" (p. xx). What is important to point out is that there is no requirement for Approach A studies to test many times before treatment. Rather Approach A, like Approach B, suggests a minimum of two to three probes and many (in fact, most) published Approach A studies in the aphasia literature do not exceed this. The authors also point out that both the Approaches A and B require a "lack of change in performance across pre-therapy" probes (p. xx).

Two questions are raised with regard to the two approaches: What is the justification for recommending two to three probes? And, what happens if a change in performance is noted during these probes? Approach A suggests a minimum (but sufficient) number of baseline probes to allow the experimenter to establish patterns of responding that can be used to predict patterns of future performance if treatment were not introduced. Hence, when fluctuating or increasing performance levels are noted during the first two to three probes, baseline probing is continued. The ultimate number of baseline probes administered, thus, is influenced by the data obtained. Conversely, with regard to Approach B, Howard et al. (2014) provide no justification for recommending two or three baseline probes or guidelines for researchers to follow if change across baseline probes is seen, that is, if performance is not stable. However, the authors do state that "while two pre-therapy probes are sufficient to provide an estimate of both level of performance and rate (slope) of change, three or more [probes] will provide a more precise estimate of both." This statement coupled with the required "lack of change in performance across pre-therapy" probes in Approach B, suggests that, as in Approach A, stable pre-treatment performance is desired.

It is also unclear why the authors say that "continuing with baseline trials … biases the baseline towards flatness, which in turn makes it more likely for the therapy to be (mistakenly) deemed effective" and "runs the risk of overestimating the effects of therapy" (p. xx). In fact, the opposite is true. Continuation of baseline to determine performance levels and slope actually safeguards against mistakenly attributing changes in performance to the treatment provided. When performance is similar across two or three baseline points, there is likely no benefit derived from continuing baseline. However, when performance is variable, that is, change in level and/or slope is seen, continuing baseline is necessary to evaluate performance patterns. Additionally, this provides researchers an opportunity to examine the source of variability. In many cases, the source of variability in people with aphasia is the aphasia itself in that day-to-day variability is a common observation. However, there may be other reasons for noted changes during baseline, including neurophysiological processes associated with spontaneous recovery or the participant may be simultaneously exposed to other treatments or environments that promote language improvement. In addition, in some cases repeated probing during baseline testing may improve performance (see later for further discussion of this). For these reasons, Approach A espouses continuing baseline to reveal any upward, or other, trends in performance before application of treatment, such that these trends are not mistaken for treatment effects.

Length of the treatment phase

Another objection raised by Howard et al. (2014)—the length of treatment—is also not specified by Approach A methodology. Again, whether treatment is provided for a

specified number of sessions (as espoused by Howard et al., 2014), or whether the length of the treatment phase is based on a pre-determined criterion is up to the experimenter and varies depending on the experimental question(s) under study. Studies that use "criterion performance" are generally interested in knowing if the treatment under investigation does or does not affect change in behaviour and aim to provide a sufficient amount of time for this to be demonstrated. Notably, these studies often use *both* a pre-determined treatment phase length as well as a response criterion. In our studies we look for criterion level performance to occur within a set time limit (e.g., 80% correct performance or a maximum of 20 treatment sessions); treatment is discontinued if criterion is reached, but continued to the phase limit if not.

One problem with training for a specified number of sessions is that not all people with aphasia learn at the same rate. For example, in our studies of sentence processing, the number of sessions required for participants to learn to produce and comprehend a set of complex sentences of a particular type varies. In Thompson et al. (2010), for example, one participant (A1) acquired target sentence types in only eight training sessions, whereas another participant (A4) required 24 sessions, the maximum number of sessions provided in that study. Indeed if the latter participant had been provided only eight sessions, little evidence of learning would have been seen. Howard et al. (2014) say that ending treatment based on participant learning curves introduces a "bias toward greater rates of improvement during therapy." Indeed, the alternative—ending a treatment phase based on a specified time period—introduces a different bias, that is, deciding that a treatment is not effective, when it is. However, neither is an Approach A requirement. Indeed, little is known about learning in aphasia, and to my knowledge, there are no studies that have systematically included length of treatment as an independent variable. This is, indeed, an empirical question.

Repeated measurement of dependent variables

The authors are correct that in Approach A single-subject-controlled experimental designs the dependent variables are repeatedly tested during the treatment phase. However, Howard et al. (2014) object to this practice because it (a) biases towards using small number of stimulus items, because testing all items is not feasible, and (b) influences performance. The issue of number of items, as discussed earlier, again is not dictated by Approach A, and there are advantages and disadvantages to including both smaller and larger stimulus sets. In addition, there are ways to arrange probe schedules during the treatment phase such that all items need not be tested during all sessions. For example, in studies conducted in my research laboratory, which are designed to examine functional relations between item sets and, hence, require many items, we create subsets of items to be used for daily (pre-session) measurement of performance, randomly assigning items from all sets to these subsets. This significantly reduces probe time, but at the same time allows inspection of learning curves during the treatment period. Another alternative is to use the multiple probe technique (Horner & Baer, 1978), an Approach A design, requiring that items from all sets be tested during the initial baseline phase of the study, but repeated testing of untrained sets is not required. This occurs only immediately prior to treatment of the untrained sets. If performance is unchanged compared to initial baseline performance, treatment is applied. If not, this constitutes generalisation across sets (when combined with a multiple baseline across participants) or a loss of experimental control (when no other design components are included).

The idea that repeated measurement influences performance is not a new one, and it can indeed be a problem. Thus, Approach A single-subject-controlled experimental designs

require safeguards against this. The multiple baseline design across item sets, for example, as discussed earlier, requires that several sets of items be tested in baseline phases, which are systematically increased in length across successive item sets (see Figure 1). If repeated measurement is responsible for changes in the dependent variable(s), this is visible on the untreated item sets and again, without additional design component, constitutes a loss of experimental control.

Analysis of the data

In the early 1960s proponents of Approach A single-subject experimentation did, in fact, oppose the use of statistics to evaluate the effects of treatment, notably for the same reasons they objected to statistical analyses of group experimental data (discussed earlier). For these, and other reasons, visual inspection of learning patterns was advanced as an alternative. However, many current Approach A, single-subject-controlled experimental researchers do use statistical methods to evaluate treatment outcomes. Indeed, both older and more recently published basic textbooks on single-subject-controlled experimental research include chapters focused on statistical analysis of the data (see Kazdin, 2003).

Still, there are difficulties with statistical analysis of single-subject experimental data. The authors discuss a few of them, including autocorrelation and other phenomenon inherent in single-subject experimentally derived data, which preclude the use of parametric statistics. These problems have been discussed at length by several researchers and are precisely the reason why statistics are not recommended for analysing such data. Howard et al. (2014) also present a discussion of the problems inherent in methods that have been deemed to be more appropriate for use with single-subject-derived data, including methods for determining effect sizes. They further present a method for comparing differences between pre- and post-treatment, analysing performance by items. Indeed, this method could augment those currently used for analysing Approach A single-subject-controlled experimental data. However, this method alone is insufficient for ruling out the effects of extraneous variables on performance (see later for further discussion of threats to internal validity).

Problems with Approach B

In addition to the aforementioned misconceptions about single-subject-controlled experimental designs (i.e., Approach A), there are several problems with Approach B. The first concerns internal validity. Although Howard et al. (2014) indicate that experimental treatment research designs need "to exclude effects from sources other than treatment" (p. x), only control "items" are used in Approach B for this purpose. There are a number of difficulties associated with using item sets for experimental control. First, experimental control is lost when control items improve during treatment of trained items. As pointed out earlier, in this case there is no way to know whether or not the treatment or some other unknown variable(s) affected change in performance. Second, there is no discussion about what constitutes control items. Are they related or unrelated to the trained items? If the items are related to one another, there is a possibility that generalisation will occur across item sets. Alternatively, if trained and control sets are unrelated, it is possible that the principles for acquisition of the two sets differ. In this case, improvement on the trained set could result from something completely unrelated to the treatment provided (i.e., extraneous variables as discussed earlier), and at the same time, a lack of improvement on the untrained set may result because whatever affected performance on the trained set

is insufficient to affect change on the control set. Indeed, a lack of change in performance during baseline followed by performance improvement when treatment is applied is not sufficient evidence for treatment effects since this does not rule out the influence of extraneous variables on changes in performance.

A third threat to internal validity in Approach B is the use of only pre- and post-treatment testing to evaluate treatment effects. This recommendation is problematic for several reasons, including regression to the mean as discussed by Howard et al. (2014), as well as day-to-day variability inherent in people with aphasia (and people with other disorders). For example, it is possible that during baseline testing, unbeknownst to the experimenter, the participant was ill and performed poorly on both treated and control items on both pre-treatment probes but was feeling quite well by the end of treatment and performed well on post-treatment probes, showing a change from pre- to post-treatment, possibly reflecting the physical health of the participant rather than the treatment provided. While seemingly absurd, experimental treatment research must be capable of addressing such a scenario (and other similar ones). With only pre- and post-treatment testing, this cannot be done. Indeed, given this case scenario, subjecting the patient's data to statistical analysis would lead to spurious results. Although comparing performance on item sets would fail to show a statistically reliable difference between sets, analyses evaluating changes from baseline to post-treatment, using the by-item method espoused by Howard et al. (2014), to examine "change in [the] probability correct" (p. xx) would show a treatment effect, when treatment was not the source of improvement. Indeed, statistical analysis of data derived from pre- and post-treatment data only obscures performance patterns.

Threats to internal validity, for example, spontaneous recovery (which can continue for more than a year post stroke) and many other factors that may influence performance cannot be ruled out. As pointed out by Howard et al. (2014) "statistics should be interpreted in combination with examination of the patterns shown in the data" (p. xx). However, when only pre- and post-treatment probes data are obtained, "trends in improvement" (p. xx) are not available for examination.

Approach B also does not address external validity. As pointed out, single-subject-controlled experimental analysis requires replication of treatment effects within and across participants to address external validity. However, there is no replication requirement in Approach B, and no other means of addressing external validity is discussed.

Finally, there are components of experimental treatment research that are overlooked by Howard et al. (2014) and are not included in their requirements for Approach B. These include reliability of measurement of the dependent measures and reliability on the independent variable, also known as treatment fidelity. Although it is beyond the scope of this paper to discuss these issues in depth, they are important components of treatment research. Reliability of measurement addresses observer bias and other issues relevant to measurement of human behaviour. Further, because the purpose of experimental treatment research is to evaluate the effects of a particular treatment on human behaviour, it behooves the experimenter to address treatment fidelity. That is, was the treatment applied correctly, as intended, or not?

Summary and conclusion

Many of the objections raised by Howard et al. (2014) to single-subject-controlled experimental analysis, referred to by the authors as Approach A, are unfounded and reflect an incomplete understanding of the requirements of Approach A designs and why

they are imposed. In addition, problems with the alternative single-case approach (Approach B) are explicated. By no means do the authors provide valid reasons for abandoning single-subject-controlled experimental research or for replacing it with alternative methods as the "way forward" in aphasia treatment research.

Note

1. Change in a process resulting from observation of that process was first noted at the Hawthorne plant of Western Electric. Production increased not as a consequence of improvements in working conditions but because management demonstrated interest in such improvements.

References

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, *12*, 199–210. doi:10.1901/jaba.1979.12-199

Connell, P. J., & Thompson, C. K. (1986). Flexibility of single subject experimental designs. Part III. Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders*, *51*, 204–214. doi:10.1044/jshd.5103.214

Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York, NY: Routledge.

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation on the multiple baseline. *Journal of Applied Behavior Analysis*, *11*, 189–196. doi:10.1901/jaba.1978.11-189

Howard, D., Best, W., & Nickels, L. (2014). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*. Advance online publication. doi:10.1080/02687038.2014.985884

Kazdin, A. (2003). *Research design in clinical psychology* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Kearns, K. P., & Thompson, C. K. (1991). Technical drift and conceptual myopia: The Merlin effect. In T. Prescott (Ed.), *Clinical aphasiology* (Vol. 19, pp. 31–41). Austin, TX: PRO-ED.

Kiran, S., Ansaldo, A., Bastiaanse, R., Cherney, L. R., Howard, D., Faroqi-Shah, Y., … Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: Standards for establishing the effects of treatment. *Neuroimage*, *76*, 428–435. PMC3552150. doi:10.1016/j.neuroimage.2012.10.011

McReynolds, L. V., & Kearns, K. P. (1986). *Single-subject experimental designs in communicative disorders*. Baltimore, MD: University Park Press.

McReynolds, L. V., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. *Journal of Speech and Hearing Disorders*, *51*, 194–203. doi:10.1044/jshd.5103.194

Schlosser, R. W. (2003). *The efficacy of augmentative and alternative communication*. Boston, MA: Academic Press.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders*, *39*, 266–291. PMC1847620. doi:10.1016/j.jcomdis.2006.02.003

Thompson, C. K., den Ouden, D. B., Bonakdarpour, B., Garibaldi, K., & Parrish, T. B. (2010). Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. *Neuropsychologia*, *48*, 3211–3227. PMC3164559. doi:10.1016/j.neuropsychologia.2010.06.036

Thompson, C. K., & Kearns, K. P. (1991). Analytical and technical directions in applied aphasia analysis: The Midas touch. In T. Prescott (Ed.), *Clinical aphasiology* (Vol. 19, pp. 41–54). Austin, TX: PRO-ED.

Thompson, C. K., Shapiro, L. P., Ballard, K. J., Jacobs, B. J., Schneider, S. S., & Tait, M. E. (1997). Training and generalized production of wh- and NP-movement structures in agrammatic aphasia. *Journal of Speech, Language, Hearing Research*, *40*, 228–244. doi:10.1044/jslhr.4002.228