

Cumulative Instructional Time and Relative Effectiveness Conclusions: Extending Research on Response Intervals, Learning, and Measurement Scale

Michelle P. Black¹ · Christopher H. Skinner^{1,2} · Bethany E. Forbes¹ ·
Merilee McCurdy¹ · Mari Beth Coleman¹ · Kristie Davis³ · Maripat Gettelfinger³

Published online: 16 February 2016
© Association for Behavior Analysis International 2016

Abstract Adapted alternating treatments designs were used to evaluate three computer-based flashcard reading interventions (1-s, 3-s, or 5-s response intervals) across two students with disabilities. When learning was plotted with cumulative instructional sessions on the horizontal axis, the session-series graphs suggest that the interventions were similarly effective. When the same data were plotted as a function of cumulative instructional seconds, time-series graphs suggest that the 1-s intervention caused the most rapid learning for one student. Discussion focuses on applied implications of comparative effectiveness studies and why measures of cumulative instructional time are needed to identify the most effective intervention(s).

- Comparative effectiveness studies may not identify the intervention which causes the most rapid learning.
- Session-series repeated measures are not the same as time-series repeated measures.
- Measuring the time students spend in each intervention (i.e., cumulative instructional seconds) allows practitioners to identify interventions that enhance learning most rapidly.
- Student time spent working under interventions is critical for drawing applied conclusions.

Keywords Session-series design · Time-series design · Learning speed · Comparative effectiveness · Response intervals

Working with students with intellectual disabilities, Yaw et al. (2014) ran two relative effectiveness studies designed to compare two computer-based flashcard reading (CFR) interventions, one with 5-s response intervals and one with 1-s response intervals. Each intervention included three trials per word, but the 1-s intervention required less student time. When word acquisition was plotted as a function of sessions, (repeated-measures session-series design with learning trials held constant), session-series figures showed both interventions resulted in similar *learning* (i.e., words acquired) per session. When word learning was measured as a function of cumulative instructional seconds (repeated-measures time-series design), time-series figures showed that the 1-s intervention resulted in superior *learning speed* (i.e., learning/time spent learning) than the 5-s intervention. These results suggested that the more rapid learning under the 1-s intervention were caused by the 1-s intervention requiring less instructional time, as opposed to 1-s trials causing more learning. Although Yaw et al. found evidence of an inverse relationship between response intervals and learning speed, the study did not provide sufficient evidence for evaluating this relationship because only two response intervals (1-s and 5-s) were applied. The current study was designed to extend the Yaw et al. research by evaluating the relationship between response intervals and learning speed using three CFR interventions (1-s, 3-s, and 5-s response intervals).

Learning is caused by events (e.g., interventions) which require time, and learning problems are typically not caused by a failure to learn, but by not learning rapidly enough (Skinner et al. 1995/2002). To prevent or remedy learning problems, comparing interventions to determine which caused

✉ Christopher H. Skinner
cskinne1@utk.edu

¹ The University of Tennessee, Knoxville, TN, USA

² EPC, University of Tennessee, BEC 535,
Knoxville, TN 37996-3452, USA

³ Lenoir City Schools, Lenoir City, TN, USA

the most behavior change or learning is insufficient; we must identify the interventions which result in the most rapid learning (Skinner 2010). Because many comparative effectiveness studies do not measure learning as a function of time spent learning, they do not allow researchers to identify which interventions cause the most rapid learning or remediation. A second purpose of this study was to contribute to the research on measurement scale by comparing relative effectiveness conclusions when learning is plotted as function of cumulative sessions and as a function of cumulative instructional seconds.

Participants were Amber and Floyd, both 10 years old and identified with a functional delay. Standardized assessment showed each had cognitive and reading achievement scores two standard deviations below the mean, but neither had significant deficits in functional or adaptive skills. Both were receiving remedial reading and math services in a self-contained classroom.

Using guidelines provided by Hopkins, Hilton, and Skinner (2011), CFR intervention and assessment programs were constructed for each participant. Both programs included words from the *Edmark Reading Program, Levels 1 & 2* (Proed 2011) in 88-point Arial style font, presented in the center of the screen, one at a time. Also, Amber's teacher selected some additional words from her curricula.

Adapted alternating treatments designs were used to compare three CFR interventions (i.e., 1-s, 3-s, and 5-s response intervals). Pre-tests took place on a Thursday and Friday. Next, experimental procedures were applied between 9:00 and 11:00 a.m. over a 2-week period at the end of the school year. The primary experimenter worked one-to-one with participants at a table in the corner of their classroom. Generally, sessions took less than 15 min.

During two pre-testing sessions, the experimenter sat with each participant and a list of 150–160 words printed down the left side of the page. The experimenter pointed to each word and prompted the student to read the word. Words read correctly within 2 s on either session were considered known and excluded. After identifying over 100 unknown words for each student, stratified (based on number of letters per word) random assignment was used to assign the 100 words to one of the four conditions, 1-s, 3-s, and 5-s CFR, and an assessment-only control condition. For each condition, the 25 words were sequenced from shortest to longest.

Sessions were conducted once per school day. Each session included 1-s, 3-s, and 5-s response-interval interventions presented in counterbalanced order across sessions. Each intervention targeted the first 10 of 25 unknown words. For each intervention, participants were instructed to click on the word START. The first word was then displayed for 1, 3, or 5 s before a recording of the word being read was played. After the recording, the word remained on the screen for 1 s before being replaced with the next word. Participants were instructed to try to read the word before the recording and repeat the word after the recording. For each intervention,

after 10 trials (one trial for each of the 10 targeted words) were completed, the program repeated the process two more times with the same words presented in random order each time. When one 30-trial intervention was completed, the same procedures were applied using the other two interventions. After the final intervention, an assessment of all 30 target words was conducted and recorded. During each assessment, each word would appear on the computer screen in random order for 2 s. If the participant read the word before it vanished from the screen, the experimenter scored it as correct. Every other session, the same procedures were used to assess the 30 target words and the first 10 control words. A word was considered acquired when it was read correctly across two consecutive assessments. Acquired words were removed from both the intervention and assessment programs and replaced with the next word assigned to that condition.

Another experimenter randomly selected and independently scored 50 % of the recorded assessment sessions by listing to the audio recording as she re-played the appropriate assessment program. Percent interobserver agreement was calculated for each session by dividing the number of agreements on words read correctly within 2 s by the number of agreements plus disagreements and multiplying by 100. Interobserver agreement for each session was 100 %.

Table 1 shows that on average, Amber acquired the same number of 1-s and 5-s words per session (2.44 words) but fewer 3-s words (1.89). When learning was assessed per instructional minute, Amber's learning speed was highest for the 1-s words (1.63 words acquired per instructional minute), which was more than twice the learning speed for the 3-s and 5-s words. Floyd's average per session data showed a clear linear pattern; he acquired the least 1-s words (0.56 per session), 0.22 more 3-s words per session (0.78 per session), and 0.44 more 5-s words per session (1.00 per session). Floyd's words acquired per instructional min across interventions were very similar, but in the opposite direction; his learning speed was highest for the 1-s words, 0.37 words acquired per instructional min, and lowest for the 5-s words, 0.29 words acquired per min of instruction.

Figure 1 displays Amber's acquisition data plotted by cumulative instructional sessions (session-series graph) and cumulative instructional seconds (time-series graph). The session-series graph shows that Amber acquired only two control words but more 1-, 3-, and 5-s words. Across the three interventions, no clear separation emerges in words learned per session. Amber's time-series graph shows a clear separation between the 1-s intervention and the other two interventions emerged. Because this separation generally grew larger as the study progressed, Amber's time-series graph provides evidence that Amber's learning speed (learning per instructional second) was superior under the 1-s intervention relative to the 3-s and 5-s interventions. There is no clear separation between the 3-s and 5-s interventions until the third session;

Table 1 Words acquired across flashcard interventions by Amber and Floyd

Students	Intervention	Mean words acquired per session	Mean words acquired per instructional minute	Total words acquired
Amber	1 s	2.44	1.63	22
	3 s	1.89	0.76	17
	5 s	2.44	0.70	22
Floyd	1 s	0.56	0.37	5
	3 s	0.78	0.31	7
	5 s	1.00	0.29	9

however, this separation did not grow larger, which suggests no consistent differences in words acquired per instructional second across the 3-s and 5-s interventions.

As with Amber, Floyd showed increases in 1-, 3-, and 5-s words acquired, while showing little improvement on control words (see Fig. 2). Although Floyd learned more 5-s words, learning per session tended to be similar across interventions following the fourth session. Also, Fig. 2 shows no differences in Floyd's word acquisition across conditions when learning was measured using cumulative instructional seconds.

Across participants, visual analysis of session-series graphs show clear increases in 1-s, 3-s, and 5-s words acquired and small increases in control words acquired, with no consistent differences in learning per session across the three interventions. These results are consistent with others who found evidence that CFR interventions can enhance word reading in

students with disabilities, but little evidence to support a functional relationship between response interval length and the amount learned per trial or session (e.g., Yaw et al. 2014).

For both students, *average* learning speed data (e.g., see Table 1) showed that as response intervals increased, learning speed decreased; however, visual analysis of Amber's time-series graphs with word learning plotted as a function of cumulative instructional seconds did not support the conclusion that the 3-s intervention resulted in *consistently* more rapid learning than the 5-s intervention. Also, visual analysis of Floyd's time-series data showed no consistent difference in learning speed across the three interventions. Therefore, the current study does not support a functional relationship between response intervals and learning speed.

Comparing relative effectiveness conclusions across session-series and time-series graphs yielded idiosyncratic

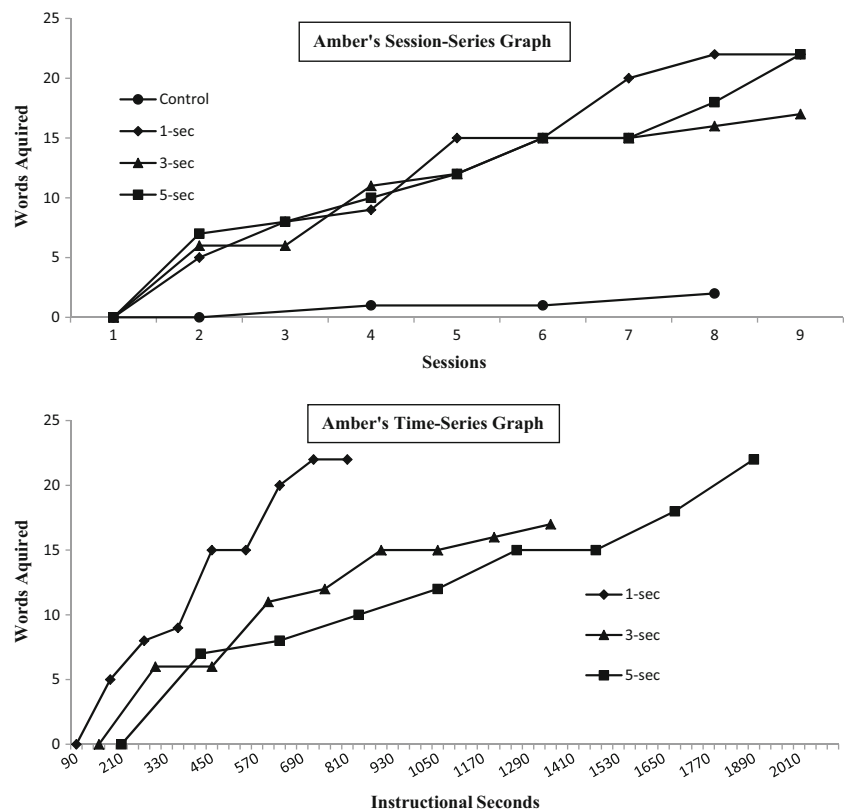
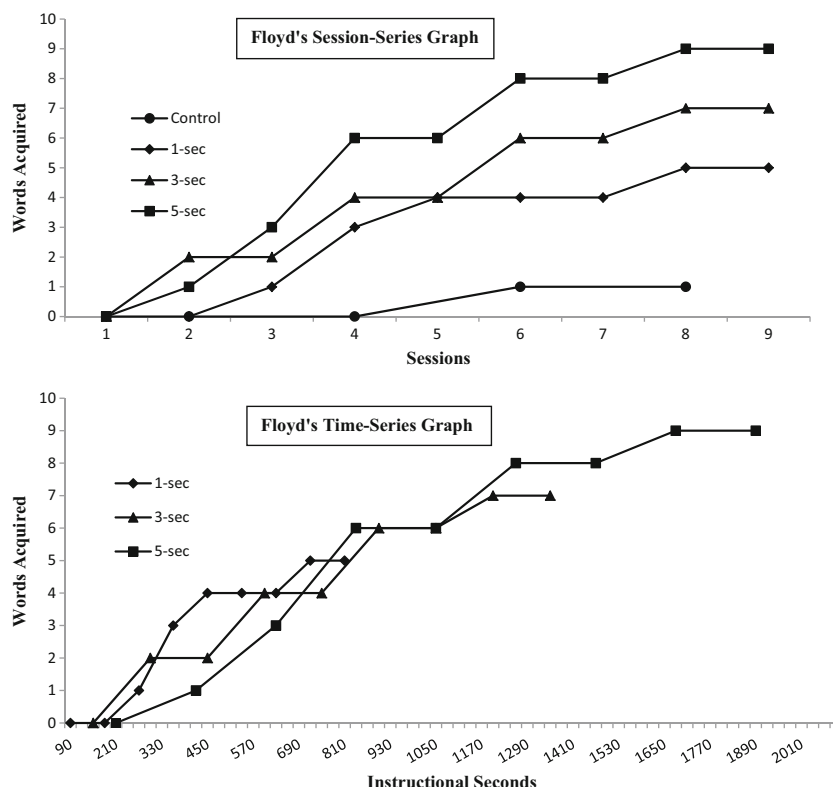
Fig. 1 Amber: The same words acquired data presented per instructional session (session-series graph) and per instructional second (time-series graph)

Fig. 2 Floyd: The same words acquired data presented per instructional session (session-series graph) and per instructional second (time-series graph)



results. For Floyd, figures plotting word acquisition as a function of cumulative instructional seconds and cumulative instructional sessions both support the same conclusions: no intervention was more effective than any other. Amber's data graphed by cumulative instructional sessions suggested that the interventions were similarly effective; however, when the same data were graphed as a function of cumulative instructional seconds, visual analysis suggests that the 1-s intervention caused larger increases in Amber's learning speed than the 3-s and 5-s interventions. These two cases demonstrate how different measurement scales (cumulative instructional session versus cumulative instructional seconds) can (e.g., Amber), but do not always (e.g., Floyd) influence our relative effectiveness conclusions. More importantly, these results suggest that Amber would benefit most from the 1-s intervention.

Researchers should address several limitations associated with our dependent variable.

We only measured word reading. Researchers should determine if CFR procedures enhance comprehension and the ability to read words when presented in connected text. Also, researchers should assess maintenance across conditions. Additionally, researchers may want to determine if voice recognition software might allow for more efficient assessment and word replacement. Finally, the idiosyncratic effects suggest finding should not be generalized to others.

In this study, learning trials were held constant per session, which may be important for answering theoretical

implications. However, when time students spend under competing interventions is not controlled, such data can mislead practitioners attempting to address applied problems (Skinner et al. 1995/2002). For example, most who care about Amber want her deficits remedied as rapidly as possible. If those people were to base their decision based on her session-series data, they may choose either the 1-s, 3-s, or 5-s intervention. However, Amber's learning speed data present in the time-series figure shows that Amber's learning speed was superior under the 1-s intervention. This idiosyncratic data suggests that the 1-s intervention will remedy Amber's deficits more rapidly than the competing interventions.

Educators attempting to remedy deficits need to know what works. But when faced with the choice of numerous empirically validated interventions (e.g., 1-s, 3-s, and 5-s CRF), theoretical studies with sessions, learning trials, or opportunities to respond held constant may not allow applied researchers to answer the next question: Which empirically validated intervention(s) remedy deficits most rapidly (Skinner 2008)? To answer this question, researchers must measure learning as a function of the time students spend learning. Thus, we encourage all researchers who run comparative effectiveness studies, including meta-analysis (see Poncy et al. 2015), to (a) evaluate learning speed using precise measures of time the student spent learning, (b) hold instructional time constant across interventions, and/or (c) clearly indicate that their evaluations of relative effectiveness should not be used to make applied recommendations (Skinner 2010). Otherwise, researchers may

encourage educators to apply theoretically supported, empirically validated procedures that are less effective at preventing and remedying deficits (i.e., take longer to remedy deficits) than other procedures (Skinner 2010).

Acknowledgments This study was completed with support from the Korn Learning, Assessments, and Social Skills Center at the University of Tennessee.

Compliance with Ethical Standards Each author declares that he/she has no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

- Hopkins, M. B., Hilton, A. N., & Skinner, C. H. (2011). Implementation guidelines: how to design a computer-based sight-word reading system using microsoft® PowerPoint®. *Journal of Evidence-Based Practices in the Schools, 12*, 219–222.
- Poncy, B. C., Solomon, G. E., Duhon, G. J., Skinner, C. H., Moore, K., & Simons, S. (2015). An analysis of learning rate and curricular scope: caution when choosing academic interventions based on aggregated outcomes. *School Psychology Review, 44*, 289–305.
- Pro-ed. (2011). *Edmark Reading Program Family of Products*. Retrieved from <http://www.proedinc.com/customer/ProductView.aspx?ID=4842>.
- Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review, 37*, 309–315.
- Skinner, C. H. (2010). Applied comparative effectiveness researchers must measure learning rates: a commentary on efficiency articles. *Psychology in the Schools, 47*, 166–172. doi:10.1002/pits.20461.
- Skinner, C. H., Belfiore, P. B., & Watson, T. S. (1995/2002). Assessing the relative effects of interventions in students with mild disabilities: Assessing instructional time. *Assessment in Rehabilitation and Exceptionality, 2*, 207–220. Reprinted (2002) in *Journal of Psychoeducational Assessment, 20*, 345–356.
- Yaw, J., Skinner, C. H., Maurer, K., Skinner, A. L., Cihak, D., Wilhoit, B., Delisle, J., & Booher, J. (2014). Measurement scale influences in the evaluation of sight-word reading interventions. *Journal of Applied Behavior Analysis, 47*, 360–379. doi:10.1002/jaba.126.