The Behavior Analyst 2004, 27, 281–286 No. 2 (Fall)

In Response

A Refocus on Response-Rate Measurement: Comment on Doughty, Chase, and O'Shields (2004)

Carl Binder
Binder Riha Associates

The recent review by Doughty, Chase, and O'Shields (2004) of studies that have examined effects of response-rate building on learning outcomes is a valuable contribution to research on use of response-rate measures for instructional decision making. It raises important experimental issues that require greater attention in the future. In its conceptual explication and analysis of published articles that meet the authors' criteria for inclusion, however, the review introduces some definitional slippage and at least one factual error that require clarification for the sake of future analysis and research.

The Great Falls Precision Teaching Project

To dispense quickly with the factual error, the authors describe Beck and Clement's (1991) report of the Great Falls Precision Teaching Project as "anecdotal evidence" (p. 7). In fact, the Great Falls project was a comparison of the Iowa Test of Basic Skills achievement scores over a 3-year period between the Sacajawea School and other elementary schools in the same district. The 20 to 40 percentile-point improvement (depending on the subtest) achieved at the Sacajawea School was attributed to 20 to 30 min per day of 1-min practices that were charted by teachers and shared with students at Sacajawea, without any other differences in curriculum or teaching methods between Sacajawea and other schools. This, one could argue, was one of the most cost-effective educational improvement demonstrations in the literature. Although it did not control for each feature of the simplified precision teaching procedures it employed, it was a robust empirical demonstration and surely not "anecdotal."

The Key Issue Is Measurement Sensitivity

At the core of the precision teaching methods and discoveries discussed by Doughty et al. (2004) is a question about measurement: Is rate of response a better predictor of important learning outcomes (e.g., retention, maintenance, application) than the more traditional percentage correct, a dimensionless quantity (Barrett, 2002; Barrett, Johnston, & Pennypacker, 1986; Johnston & Pennypacker, 1980)?

In education, where percentage correct is ubiquitous, we sometimes forget Skinner's assertion that his most significant contributions were "the use of rate of responding as a basic datum and the so-called cumulative record" (Evans, 1968, p. 103). Vargas (1977) clarified the implications of this statement for instructional applications when she said,

> Teaching ... is not only producing new behavior, it is also changing the likelihood that a student *will* respond in a certain way. Since we cannot see likelihood, we look instead at how frequently a student does something. We see how fast he can add. The student who does problems

Address correspondence to Carl Binder, Binder Riha Associates, 4966 Wilshire Dr., Santa Rosa, California 95404 (e-mail: CarlBinder@aol.com). For downloadable versions of many publications cited in this article visit www.fluency.org.

correctly at a higher rate is said to know addition facts better than one who does them at a lower rate. (p. 62)

In the instructional context, a measure of response likelihood, strength, or probability refers to the *ability* of an individual to behave in a particular (skillful) way and not merely to that individual's *inclination* to do so. Thus, use of response-rate measures is essential at every step in controlled research that seeks to verify clinical discoveries from precision teaching. A possibly more interesting way to frame rate-building research than as a comparison of the effects of controlled trials with the effects of self-paced practice on retention, application, and other learning outcomes is to ask if rate of freely emitted responding better predicts these learning outcomes, whether that rate is the product of controlled-trials practice, self-paced practice, or some combination. Does it tell us more than percentage correct?

Achieving Normal Competent Rates of Response—Not "High" Rates

In their review, Doughty et al. (2004) frequently use the term *high* when referring to response rates achieved with precision teaching. This modifier unnecessarily biases the discussion, because it suggests rates that are out of the ordinary, extreme, or unusual. As Barrett (1979) illustrated with behavioral education for developmentally delayed students aimed at helping them perform in so-called normal environments, response-rate measures enable us to detect clear differences in ability among competent adults, children, and disabled learners and lead to better instructional decisions than do percentage correct scores, which do not reveal such differences. Her data (reprinted in Binder, 2003; Johnson & Layng, 1992) show systematic differences in rates of correct responding among the three aforementioned populations that are not reflected in percentage correct scores. With these data in mind, the phrase "high response rates" should be changed in Doughty et al. to "normal ranges of response rate." There is nothing unusual or extreme about using a more sensitive measure (i.e., response rate) to identify normal ranges of competence and then applying procedures that enable learners to achieve those normal ranges. To the contrary, this is what Barrett called "communitization" or behavioral normalization.

Defining Application

When describing key learning outcomes cited in the precision teaching literature, Doughty et al. (2004) redefine *application* in a way that differs from how the term is used in that literature. When Starlin (1971), Haughton (1972), and their colleagues first determined that students who could correctly perform key component skills (e.g., adding numbers) at or above certain response rates (e.g., 40 to 50 per minute) were successful in more advanced composite skills (e.g., math story problems) while those with lower rates (e.g., 20 to 30 per minute) were not, it dawned on them that combination of smaller units of behavior (components) into larger units (composites) might depend on development of higher component response rates. Clinical measures and observations since the early 1970s have confirmed this result (Binder, 2003), as have a few published experimental studies (e.g., Van Houten, 1980). At one point referred to as *compounding* (Haughton, 1980), this was the origin of the term *application* later communicated with Haughton's acronyms R/APS (retention/application performance standards) and REAPS; these are further discussed by Johnson and Layng (1992, 1996), Binder (1993, 1996), and others.

Despite this documented history, Doughty et al. (2004) redefine *application* as "generalization, or the occurrence of a skill in novel stimulus conditions" (p. 9). Although generalization might in some instances be a result of achieving competent rates of responding, a narrower generalization gradient might just as well occur, depending on the specificity of stimulus control during rate building. In any case, Doughty et al.'s definition of *application* does not correspond to that of the precision teaching and fluency literature, and therefore does not reflect claims made or questions asked among precision teaching practitioners and researchers. Instead, precision teachers are looking for application when they check to see if learners readily combine components into composites (also called compounds) during the instruction and practice of behavior composites once they have achieved specified ranges of count-per-minute responding on those components (Fabrizio & Moors, 2003). Bucklin, Dickinson, and Brethower (2000), one of the studies reviewed by Doughty et al., demonstrated this application effect in controlled research with a stimulus equivalence task. When precision teachers refer to *adduction* (Binder, 1996; Johnson & Layng, 1996), they are pointing to a special case of application in which behavior components at certain response rates combine with no explicit training on the composites. Optimizing adduction is a key element of generative instruction (Johnson & Layng, 1992; Johnson & Street, 2004).

REAPS and Rate of Response

In addition to a different definition of *application*, Doughty et al. (2004) introduce an error regarding Haughton's acronym REAPS (Binder, 1996; Johnson & Layng, 1996). Printed incorrectly as REAPs (small s) in their review, the authors claim that the acronym represents possible learning outcomes listed as "retention, endurance, application, and performance standards" (p. 8). In fact, the acronym posed a challenge from Haughton to identify performance standards (as count-per-minute ranges) that optimize retention, endurance, and application. This was an empirical alternative to the methodology of collecting samples of response rates from groups judged to be competent by other means, including social validation (Barrett, 1979; Haughton, 1980; Wood, Burke, Kunzelman, & Koenig, 1978).

Like reading teachers and language specialists, precision teachers use the term *fluency* to point to the temporal dimension of competent performance. Reading teachers measure fluency as count of words read per minute. Thesaurus entries such as *fluidity, smoothness, quickness, dexterity,* and *rapidity* reflect both the qualitative and temporal features of speech or other behavior referenced in everyday use of the term. Defining the term *fluency* separate from the time dimension, which Doughty et al. (2004) seem to do in their review (i.e., by appealing to this author and to others who have emphasized retention, endurance, and application as outcomes associated with achieving competent response rates), is potentially confusing. To define fluency as a level of performance that predicts retention, endurance and application without specifying a frequency range within which those outcomes are optimized is not consistent with use of the term by precision teachers. It ignores the fact that percentage correct—the available alternative—cannot distinguish between levels of performance that do and do not predict those outcomes.

Again, the central question is not whether practice procedures carried out beyond the point of 100% accuracy in the absence of time-based measurement can result in better retention, application, and so on. Extensive overlearning research during the 20th century demonstrated that such procedures do result in better long-term outcomes (Binder, 1993). The more fundamental issue in evaluating precision teaching is whether using Skinner's response-rate measure rather than percentage correct during teaching and learning is a more sensitive way to distinguish competence from incompetence and to make instructional decisions. In fact, precision teachers use a variety of teaching and practice procedures, some timed and others untimed, to produce normal rates of freely emitted behavior. What distinguishes precision teachers from conventional educators (including other behavioral educators) is their use of daily response-rate measures graphed on the standard celeration chart (Pennypacker, Gutierrez, & Lindsley, 2003) for making instructional decisions. It is how they measure the effects of their procedures, not the myriad of procedures themselves, that primarily differentiates them from more conventional educational practitioners. Although Doughty et al. (2004) do not explicitly address this more general issue, reframing the discussion with this in mind may help to prevent future misunderstandings about precision teaching and its foundation in Skinner's response-rate measure.

Defining Endurance and Distractibility

In precision teaching research and practice, the term *endurance* has referred from its earliest use to a combination of features that characterize truly masterful fluent behavior (Binder, 1996). Masterful performers are able to continue performing over comparatively extended periods of time without being unusually subject to errors or distraction from events that can potentially compete for stimulus control. Although it is possible to observe unstable behavior under a variety of conditions—for example, in the presence of conditioned aversive stimuli or with extinction procedures—the type of stability referred to by precision teachers (e.g., Johnson & Layng, 1996) is a feature of the overall "endurance effect" observed in classrooms (Binder, Haughton, & Van Eyk, 1990).

Early laboratory investigation of the endurance effect (Binder, 1996) used a free-operant analogue of the discrete-trials procedure reported by LaBerge and Samuels (1974) in a verbal learning paired-associate study. They observed the effects of a potentially competing stimulus introduced just prior to a paired-associate trial as a distraction, and measured the effect on latency and accuracy of response. In pilot experiments using free-operant procedures reported by Binder, the distracting stimulus was competing directly for stimulus control when the subject heard numbers read through earphones while he or she attempted to complete math calculations by responding vocally to problems on a practice sheet.

Although the use of ambient noise to produce unstable responding has slipped into evaluations of stability conducted by precision teachers, original use of the term *distractibility* referred to the influence of potentially competing stimulus control, suggesting that behavior at rates closer to normal competent ranges (i.e., of greater response strength) might compete more effectively with potentially conflicting stimulus control than behavior at lower rates or response strength. This notion of competing stimulus control offers a more functionally defined and perhaps more interesting conceptual underpinning for both research and application concerning the effects of rate building on distractibility. Because Doughty et al. (2004) did not specifically acknowledge this distinction, it is important to mention it in relation to their review and use of the term *stability*; the difference between general ambient noise and directly competing stimulus control has implications for experimental designs that assess distractibility.

Measuring Rate of Response after Control Procedures

One of the most important topics discussed by Doughty et al. (2004) is measurement of the relative effects of various kinds of control procedures that precision teaching researchers should consider including in their designs. They specifically recommend "teaching some skills in a rate-building format while teaching others in a rate-controlled format" (p. 20). Stated another way, this is a recommendation to compare the effects of self-paced practice procedures that include count-per-minute goals (rate building) with rate-controlled trials procedures that provide the same number of response opportunities, total exposure time, or rate of reinforcement.

Typically, those who conduct such research control for amount of practice by providing the same number of practice trials as the number of responses in a yoked procedure in which learners move at their own pace. Controlling for total time of exposure yokes time during rate-controlled trials to total time in which learners respond at their own pace. Controlling for amount of reinforcement might likewise compare the effects of rate-controlled trials with the effects of practicing freely emitted behavior in one or more control conditions, although number of reinforcements (one per session or one per response?) and sources of reinforcement (extrinsic, or both extrinsic and intrinsic?) can be difficult to define unambiguously in procedures that allow continuous self-paced responding.

One could argue from a technology standpoint that if a given number of rate-controlled trials were to produce equivalent retention, application, or other learning outcomes as the same number of freely emitted responses during self-paced practice, the less time-consuming procedure (i.e., freely emitted responses) is more efficient and therefore practically a superior procedure. But for those who wish to control for amount of practice, number of reinforcers, or any other features using controlled trials in comparison with self-paced practice procedures, it is critical to probe freely emitted response rates after the controlled trials conditions and before tests for learning outcomes such as retention and application to compare the immediate and direct impact of each procedure on response strength, using Skinner's measure.

An obscure finding in the overlearning literature (Binder, 1993) is relevant here. Various researchers (e.g., Kruger, 1929) found that the number of trials required to achieve a 100% criterion in controlled trials procedures does not predict the number of trials beyond 100% accuracy required to achieve a given level of retention. In other words, there is no way to predict for any individual how many additional trials will be needed to achieve a specific degree of the longer term outcome. On the other hand, precision teachers have routinely found that achieving different count per minute levels of performance predicts different degrees of retention, application, and so on. Although there is no doubt that controlled trials do affect these outcomes, the measurement ceiling defined by 100% correct and the artificial limit on response rate imposed by the trials procedures themselves leave no way of determining how much practice should be added beyond 100% correct to achieve the desired outcomes. If there is a measurement that would allow us to predict these outcomes, it seems most likely to be Skinner's rate of response. Merely shifting from rate-controlled trials procedures and materials to procedures that allow learners to respond as quickly as they are able can double or triple response rates without any other intervention (Binder, 2003); thus, failing to probe response rates after rate-controlled procedures leaves out an essential piece of the puzzle.

As a technical matter, preventing learners from responding at their own pace (i.e., discrete trials) can reduce the likelihood of their immediately adapting to self-paced procedures by responding freely and repeatedly. This phenomenon is evident in students with autism who often have difficulty learning to "do the next one" after histories of rigorously controlled discrete trials. For that reason, rate probes after discrete trials should probably include three or more timings to allow adaptation to the self-paced procedure, sometimes called *warm up*. In this author's experience, the most common error in graduate students' experimental designs for fluency research is the failure to include response-rate probes at all appropriate points in a sequence of procedures. Perhaps this is to be expected, because it is the use of response rates to assess the effects of instruction that is so new and foreign for most of us who have literally grown up in a percentage correct world.

Conclusion

Doughty et al. (2004) have served the field by raising a number of experimental design issues to a new level of visibility. This article is an attempt to clarify potential confusions in their review. It also seeks to introduce several additional considerations that should be integrated into any research plan on the use of response-rate measures to define competence or to make instructional decisions.

REFERENCES

Barrett, B. H. (1979). Communitization and the measured message of normal behavior. In R. York & E. Edgar (Eds.), *Teaching the severely handicapped* (Vol. 4, pp. 301–318). Columbus, OH: Special Press.

Barrett, B. H. (2002). *The technology of teaching revisited: A reader's companion to B. F. Skinner's book.* Concord, MA: Cambridge Center for Behavioral Studies.

Barrett, B. H., Johnston, J. M., Pennypacker, H. S. (1986). Behavior: Its units, dimensions, and measurement. In R. O. Nelson & S. C. Hayes (Eds.), *Conceptual foundations of behavioral assessment* (pp. 156–200). New York: Guilford.

Beck, R., & Clement, R. (1991). The Great Falls precision teaching project: An historical examination. *Journal of Precision Teaching, 8(2),* 8–12.

Binder, C. (1993). Behavioral fluency: A new paradigm. *Educational Technology, 33(10),* 8–14.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19,* 163–197.

Binder, C. (2003). Doesn't everybody need fluency? *Performance Improvement, 42(3),* 14–20.

Binder, C., Haughton, E., & Van Eyk, D. (1990). Increasing endurance by building fluency: Precision teaching attention span. *Teaching Exceptional Children, 22(3),* 24–27.

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13(3),* 140–163.

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst, 27,* 7–23.

Evans, R. I. (1968). *B. F. Skinner: The man and his ideas.* New York: Dutton.

Fabrizio, M. A., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4,* 23–36.

Haughton, E. C. (1972). Aims: Growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing* (pp. 20–39). Arlington, VA: Council on Exceptional Children.

Haughton, E. C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching, 1(3),* 3–20.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47,* 1475–1490.

Johnson, K. R., & Layng, T. V. J. (1996). Fluency. *The Behavior Analyst, 19,* 281–288.

Johnson, K. R., & Street, E. M. (2004). *The Morningside model of generative instruction: What it means to leave no child behind.* Concord, MA: Cambridge Center for Behavioral Studies.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Erlbaum.

Kruger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12,* 71–78.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Pennypacker, H. S., Gutierrez, A. Jr., & Lindsley, O. R. (2003). *Handbook of the standard celeration chart.* Concord, MA: Cambridge Center for Behavioral Studies.

Starlin, C. (1971). Evaluating progress toward reading proficiency. In B. Bateman (Ed.), *Learning disorders: Vol. 4. Reading* (pp. 389–465). Seattle, WA: Special Child Publications.

Van Houten, R. (1980). Study demonstrating the effects of building rate of writing answers to simple multiplication facts on the performance of long multiplication. In R. Van Houten (Ed.), *Learning through feedback: A systematic approach for improving academic performance* (pp. 24–25). New York: Human Sciences Press.

Vargas, J. S. (1977). *Behavioral psychology for teachers.* New York: Harper and Row.

Wood, S., Burke, L., Kunzelman, H., & Koenig, C. (1978). Functional criteria in basic math proficiency. *Journal of Special Education Technology, 2,* 29–36.