Focus on Response Rate Is Important but not Sufficient: A Reply

Philip N. Chase

West Virginia University

Shannon S. Doughty

Parsons State Hospital and Training Center

Elizabethann O'Shields

West Virginia University

When we heard that *The Behavior Analyst* was publishing a response (Binder, 2004) that was critical of our review of rate-building procedures (Doughty, Chase, & O'Shields, 2004), we feared that perhaps we had missed something vital. We looked forward to reading the response, however, because our primary purpose in writing the review was to stimulate scholarship on precision teaching and its relation to behavioral education. Part of that scholarship is opening the search for experimental validation of the procedures used and recommended by precision teachers for rate building. Thus, we sincerely hoped that someone would point out studies we had missed in our review that would demonstrate the success of these procedures for producing important educational outcomes. We imagined that if anyone would have access to such studies, it would be someone like Carl Binder, who has studied precision teaching for many years.

Although Binder (2004) did not identify studies that have experimentally validated the use of rate-building procedures, he made some points that might advance the field. Binder provided definitions of learning outcomes (often referred to as fluency outcomes) and repeated the emphasis of Doughty et al. (2004) to separate the issues of

measurement from those of procedure. Binder also called for a renewed emphasis on rate of responding that predicts learning outcomes rather than simply high rates. For the most part, however, we were disappointed in Binder's reply because we believe it muddled the focus of our review: an examination of the *experimental* evidence for the procedures used and suggested by the field of precision teaching. Binder did not provide examples of experimental evidence; instead he provided a lucid argument from an authority on precision teaching.

Such argument from authority is exactly what motivated us to write the review in the first place. We went looking for answers in the literature to problems our students and research participants have encountered when learning new skills. What we found was a paucity of experimental evidence and quite a few arguments from authority. Binder's response to our review was another example of these arguments. Using words and phrases such as "refocus" in the title, "at least one factual error" and "definition slippage" in the opening paragraph, "more interesting" research questions (p. 282), and "the errors of graduate students" in the concluding paragraph are prime illustrations of the authoritative tone of the article. The current authors take it that we, as well as readers, should simply accept these comments because they were put forth forcefully by an authority. But that is

Address correspondence to Philip N. Chase, Psychology Department, P.O. Box 6040, West Virginia University, Morgantown, West Virginia 26506 (e-mail: pchase@wvu.edu).

the antithesis of science and evidence-based decision making.

Our goal for the review was to provide readers with the evidence and to suggest further experimentation. In this light, we would like to comment on a few of the problems Binder (2004) found with the review so that future research can be encouraged on these important educational issues. Our comments focus on points made by Binder that we found confusing or misleading. In addition, since we began our response to Binder, another article has been published concerning our review (Kubina, 2005). Because that article repeats some of Binder's themes, we also comment on them here. Throughout, however, we describe points of agreement and end on a constructive note.

Factual Error

We were first struck by the argument that our inclusion of Beck and Clement (1991) as an example of an anecdotal report was a "factual" error (Binder, 2004). Binder pointed out that the Great Falls Precision Teaching Project described by Beck and Clement was a "robust empirical demonstration and surely not anecdotal" (p. 281). There are many good things that can be said about the Great Falls Precision Teaching Project as reported by Beck and Clement: it was empirical and led to citations and honors for excellent work to help children learn. The Great Falls Precision Teaching Project is one of the examples of school-based programs, along with Morningside Academy, that "have reported extraordinary success in the acquisition and maintenance of academic behavior by their students" (Doughty et al., 2004, p. 10). It is important to praise good behavior-analytic work on educational problems as often as possible; therefore, we are glad Binder does so.

To say that describing Beck and Clement (1991) as anecdotal is a factual error, however, goes too far. The part of the review in which we referenced Beck and Clement was in the

presentation of the need for an evaluative review. We had found articles that were either anecdotal or conceptual, but we had found none that provided an evaluative review. To be perfectly clear, we were not referring to the Great Falls Precision Teaching Project in the sentence referring to these general articles; we used that specific article as an example of the kind of anecdotal articles we had found in the literature.

According to both the *Oxford English Dictionary* (2005) and *Webster's New Collegiate Dictionary* (1967), the first meaning of *anecdote* is that it refers to an unpublished document. This definition was clearly not the one to which we were referring, as we cited a publication. The second definition is that an anecdote is a narrative of an incident told as being interesting or striking. This was how the word was used in the review. Beck and Clement's (1991) article is a description of an interesting event. The interesting event is a school-based project that refers to evidence but does not provide experimental evidence, because it is not detailed enough for someone to either evaluate or replicate what was done on the basis of the article.

Perhaps other peer-reviewed publications reporting on the Great Falls Precision Teaching Project include the details necessary to evaluate or replicate their experimental evaluation of precision teaching. If these reports can be examined for experimental evidence for replicable procedures that affect important educational outcomes, then they should be. If we missed these reports, they should be pointed out and the results described. There are two problems with this recommendation, however. First, Beck and Clement (1991) provided only one reference, and it was not to a report on the experiment conducted. We could not find a published, peer-reviewed journal article on the experiments conducted. Second, as Binder (2004) indicated, the experiments conducted on the Great Falls Project did not separate the ef-

fects of different procedures used. This fact puts the study into the same category as many articles we did review: They did not control for other plausibly relevant variables.

We believe our characterization of the article as anecdotal still stands. Our intent, as should be clear from reading the whole paragraph in context, is that we were distinguishing between research reports, conceptual articles, and reviews of research. We were justifying a review of published reports of research for which the data presented could be evaluated and the procedures could be potentially replicated. Clearly Beck and Clement (1991) is not such an article.

Refocus on Measuring Rate

Our second reaction was to the implication that the review had somehow failed to make the distinction between rate of responding as a measure and rate-building procedures. We agree with Binder, however, and stated explicitly and early in our review (see *Defining and Clarifying Terms and Methods*, p. 8) that it is important to distinguish between the use of rate of responding as a measure and the use of procedures that have been derived from precision teaching. This distinction also is supported by others who have investigated precision teaching. For example, Barrett (2002) stated that in addition to its focus on rate of responding, "precision teaching spawned a number of methods: brief timings, increasing the frequency of prerequisite skills, determining frequency-based standards, etc." (p. 74). We think this is a useful distinction that we made clear in the review; we then reviewed the evidence for whether one of these procedures, rate building, leads to changes in important measures of academic performance.

The difference between the review's presentation of this distinction and Binder's, however, is that Binder (2004) treated measuring rate of responding as taking precedence over ex-

amining procedures. This treatment seems clear, in that Binder repeatedly stated alternative questions about the use of rate of response as a predictor. Such priority suggests that behavior analysts' most important contributions are studying correlations between rate of responding and other measures of academic, educational, and clinical accomplishment. This stance is curious for a behavior analyst even if it can be attributed to B. F. Skinner (Binder, 2004). Certainly measures are important, and certainly correlations are important, but both are most important for finding similarities and differences among the kinds of manipulations of the environment that can be made. As behavior analysts, we are interested in more than prediction; we are interested in experimental control. Thus, we think there is a strong place for the evaluation of claims that procedures, curriculum, and so on, affect student achievements.

Definitional Slippage

Binder (2004) spent much of the article arguing for particular definitions of learning outcomes. Kubina (2005) also addressed the definitions of learning outcomes used in Doughty et al. (2004). Both Binder and Kubina use strong language to suggest that the definitions in the review were incorrect (e.g., "definitional slippage," Binder, p. 281; "misclassified," Kubina, p. 75). The arguments also imply that the alternative definitions would make a difference to the findings reported. Neither Binder nor Kubina, however, provides evidence that experiments using their preferred definitions invalidated the findings of the review. They simply claimed that previous articles had defined *application* and *endurance* differently. Kubina refers to two articles that were not available at the time of the review (Kubina, Young, & Kilwein, 2004; Lin & Kubina, in press). We hope these articles shed some further experimental light on the relations

between rate of response and learning outcomes.

Three further points can be made about the “correct” definitions of learning outcomes. First, one of our empirical findings, stated on page 11 both in the text and in a footnote, was that the inconsistent and infrequent use of any of the outcomes that we characterized as testing for generality, including application, suggested that we use Skinner’s term *extension* (Skinner, 1957). Second, if one examines our definition of extension (p. 11), one can see that it subsumed the definition provided by Binder (2004) and Kubina (2005). As a consequence, our review was likely to more inclusive, not less. We might add that our review attempted to be as inclusive as possible in finding experiments reported in peer-reviewed journals that supported the relation between rate-building procedures and learning outcomes. This is why we included Miller, Hall, and Howard (1995) as a study of endurance even though they did not themselves categorize it as such (see Kubina, p. 74, for an alternative view).

Third, we qualified our definitions of learning outcomes very clearly with the sentence: “These outcomes have been discussed in detail elsewhere (e.g., Binder, 1996; Johnson & Layng, 1996; Weiss, 2001) and *are defined only briefly here* [italics added]” (p. 8). It was in these other articles that the definitions were provided, and even then, the definitions and names of these outcomes were not consistent. To spend time on the evolution of these terms and the priority of one definition over another would have detracted from our purpose and, more important, would not have changed the conclusions of the review. To claim that these terms are technical, as Kubina (2005) states, is not supported by the literature we reviewed.

We think the general thrust of defining success in terms of learning outcomes, however, is important. In fact, our purpose in using learning outcomes as criteria in the first place was because

we agree that these kinds of outcomes are exactly what teachers, experimenters, and other educators should be focusing on achieving. As described by the first author of the present paper in a recent chapter on behavioral education,

Many teachers state that they are interested in teaching behavior that is maintained over time, that endures under situations that are less than ideal, that occurs under a range of appropriate situations, and that can be combined with other behavior when the need arises. Achieving these outcomes is a behavioral definition of fluency (Johnson & Layng, 1992). For example, fluent sailors *maintain* their knowledge of rigging from one summer to the next. Their rigging skills *endure* even when the wind is high. They are able to *apply* their rigging skills to a variety of boats. And they can *combine* their rigging skills with their knot-tying skills when a piece of equipment breaks. These outcomes are based on behavior-analytic ties to selectionism (i.e., selecting repertoires that are adaptive for the individual, see Johnson & Layng, 1992). The outcomes also specify pragmatic criteria for testing whether instruction is successful. (Chase, 2003, p. 354)

To argue that these criteria have somehow slipped or were misrepresented in the review seems contrary to points made in the review and by the authors elsewhere.

Practice versus Rate Criteria

Another claim Binder (2004) made was that studies of practice have often been criticized because they do not lead to a specification of amount of practice. In defense of this criticism, Binder cited a finding from the overtraining, overlearning, and automaticity literatures: The number of trials beyond 100% accuracy that is related to other measures of performance is affected by individual differences. Thus, one individual might benefit from hundreds of trials, and another individual might benefit from a few extra practice trials. The same argument, though, can be leveled at rate of responding. Rate of responding varies considerably across individuals who perform well on outcome measures (cf. Johnson & Layng, 1992, Figure 3). In fact, Binder made the same argument when he referred to the range of rates that might

predict these outcomes (p. 283). We agree that ranges of any measure are likely to be important to specify, but to say that a range of rate measures is better than a range of trial measures (i.e., practice) is an empirical generalization that is not supported by the literature. We think it is also important to recognize that percentage correct is not the only alternative measure to rate of responding that might be useful (see Binder, pp. 281–282). Measuring behavior in time can also include speed of responding or reaction time. We are proposing that whether any measure has priority needs to be discovered through research.

Matters of Efficiency

Binder (2004) and Kubina (2005) also argued that even if practice is the critical component of rate-building procedures, the procedures involved in 1-min timings are a more efficient way to achieve practice. Although we made a similar statement in the review (p. 18), we went on to state that more research needs to be done before drawing this conclusion. Before claiming that rate-building procedures are more efficient, it is important to test the equally plausible alternative that students do not have to achieve criterion response rates to achieve the learning outcomes measured. It is plausible that a more efficient procedure would be to have students practice a few extra times beyond a mastery criterion or to have the students go through a better designed and empirically tested program of instruction rather than engaging in extra practice or rate building.

Again, like the issue of what constitutes a good predictor, there are many individual differences in skills that need to be repeated within an individual and across individuals. Thus, the jury is still out on the relative efficiency of rate-building procedures. Data from well-controlled studies are needed before we can conclude that rate building is more efficient than other kinds of procedures.

High Rates

On a constructive note, we applaud Binder (2004) for recognizing that the use of the term *high rates of behavior* is probably good to drop from the vocabulary of precision teaching. Of course the term *high* is relative, and if one is investigating the differences between procedures that produce different rates, it is perfectly acceptable to call the one that is higher “a high rate” and the one that is lower “a low rate.” In addition, we wish to point out that we are not the only ones in the precision-teaching literature who have used the term “high rates” or “maximum rates” or “responding as quickly as you can,” all of which suggest some kind of “rates out of the ordinary” (Binder, p. 282). We found examples from at least seven precision teaching articles using these phrases (e.g., Beck & Clement, 1991, throughout; Berquam, 1985, p. 322; Evans, Mercer, & Evans, 1983, p. 29; Johnson & Layng, 1992, p. 1485; Van Houten, Morrison, Jarvis, & McDonald, 1974, p. 547; Weiss, 2001, p. 185). But we think that Binder focused on an important issue. The purpose of rate building should not be to create absurdly high rates of behavior or high rates of absurd behavior, but rather to produce rates of responding that predict the kinds of learning outcomes we have discussed. How these rates are determined is an empirical issue for precision teachers and experimenters who are interested in precision teaching. Whether to use expert rates, adult rates, or rates of other students has not been resolved. Binder and others have argued that the rates must have some empirical basis in their relation to other outcomes, like retention, endurance, and application. And we agree. We hope others help by reporting these rates when they are found to be predictive of these outcomes. Moreover, we hope that others will go beyond correlation to determine the procedures that control these all-important educational outcomes.

Conclusion

Although there are some subtle points made by Binder (2004) that are worth repeating, we fear that Binder and the subsequent response by Kubina (2005) will be read and cited as successfully rebutting our criticism of rate-building procedures. We disagree. We have yet to read a peer-reviewed article that experimentally demonstrates an unambiguous relation between rate-building procedures and the kinds of learning outcomes discussed here and elsewhere. The evidence seems clear to us: Precision teaching in its entirety, like other forms of behavioral instruction (e.g., programmed instruction, direct instruction, the personalized system of instruction), is effective. Equally evident is that rate of responding is an effective measure of performance for some kinds of behavior. The contribution of components that are often specified as distinguishing precision teaching from other forms of successful behavioral instruction (e.g., rate-building procedures) is less clear. Whether rate of responding is always the preferred measure of behavior also is less clear. Other measures that include time as a dimension of behavior, such as speed scores, reaction times, and latencies, may be preferred under some circumstances. We continue to study these potential contributions experimentally and to use rate of responding and other measures of behavior in time, and we hope others will as well.

REFERENCES

- 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.
- Beck, R., & Clement, R. (1991). The Great Falls Precision Teaching Project: An historical examination. *Journal of Precision Teaching*, 8, 8–12.
- Berquam, S. (1985). Precision teaching: Does it work? *B. C. Journal of Special Education*, 9, 321–327.
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19, 163–197.
- Binder, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields. *The Behavior Analyst*, 27, 281–286.
- Chase, P. N. (2003). Behavioral education: Pragmatic answers to questions of novelty and efficiency. In K. A. Lattal & P. N. Chase (Eds.), *Behavior theory and philosophy* (pp. 347–367). New York: Plenum/Kluwer.
- 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, S. S., Mercer, C. D., & Evans, W. H. (1983). The relationship of frequency to subsequent skill acquisition. *Journal of Precision Teaching*, 4, 28–34.
- Oxford English Dictionary*. (2005). Oxford, UK: Oxford University Press. [electronic version]
- 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). On terms and procedures: Fluency. *The Behavior Analyst*, 19, 281–288.
- Kubina, R. M. (2005). The relations among fluency, rate building, and practice: A response to Doughty, Chase, and O'Shields (2004). *The Behavior Analyst*, 28, 73–76.
- Kubina, R. M., Young, A. E., & Kilwein, M. (2004). Examining an effect of fluency: Application of oral word segmentation and letters sounds for spelling. *Learning Disabilities: A Multidisciplinary Journal*, 13, 17–23.
- Lin, F. Y., & Kubina, R. M. (in press). The relationship between fluency and application for multiplication. *Journal of Behavioral Education*.
- Miller, A. D., Hall, S. W., & Heward, W. L. (1995). Effects of sequential 1-minute time trials with and without inter-trial feedback and self-correction on general and special education students' fluency and math facts. *Journal of Behavioral Education*, 5, 319–345.
- Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley.
- Van Houten, R., Morrison, E., Jarvis, R., & McDonald, M. (1974). The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis*, 7, 547–555.
- Webster's seventh new collegiate dictionary*. (1967). Springfield, MA: Merriam.
- Weiss, M. J. (2001). Expanding ABA interventions in intensive programs for children with autism: The inclusion of natural environment training and fluency based instruction. *The Behavior Analyst Today*, 2, 182–185. Retrieved January 18, 2004, from <http://www.behavior-analyst-online.org/BAT/BAT23.pdf>